Reclaiming the Coast Salish woolly dog pp. 1236 & 1303

The future of immunotherapies for Alzheimer's disease p. 1242

All-day thermoregulatory clothing pp. 1247 & 1291

Science

$15
15 DECEMBER 2023
SPECIAL ISSUE
science.org

AAAS

2023

BREAKTHROUGH OF THE YEAR

OBESITY MEETS ITS MATCH

Blockbuster weight loss drugs show promise for a wider range of health benefits

By **Jennifer Couzin-Frankel**

Obesity plays out as a private struggle and a public health crisis. In the United States, about 70% of adults are affected by excess weight, and in Europe that number is more than half. The stigma against fat can be crushing; its risks, life-threatening. Defined as a body mass index of at least 30, obesity is thought to power type 2 diabetes, heart disease, arthritis, fatty liver disease, and certain cancers.

Yet drug treatments for obesity have a sorry past, one often intertwined with social pressure to lose weight and the widespread belief that excess weight reflects weak willpower. From "rainbow diet pills" packed with amphetamines and diuretics that were marketed to women beginning in the 1940s, to the 1990s rise and fall of fen-phen, which triggered catastrophic heart and lung conditions, history is beset by failures to find safe, successful weight loss drugs.

But now, a new class of therapies is breaking the mold, and there's a groundswell of hope that they may dent rates of obesity and interlinked chronic diseases. The drugs mimic a gut hormone called glucagon-like peptide-1 (GLP-1), and they are reshaping medicine, popular culture, and even global stock markets in ways both electrifying and discomfiting. Originally developed for diabetes, these GLP-1 receptor agonists induce significant weight loss, with mostly manageable side effects. This year, clinical trials found that they also cut symptoms of heart failure and the risk of heart attacks and strokes, the most compelling evidence yet that the drugs have major benefits beyond weight loss itself. For these reasons, *Science* has named GLP-1 drugs the Breakthrough of the Year.

In honoring these therapies, we also acknowledge the uncertainties, even anxieties, this sea change brings. We recognize, too, that obesity comes with medical and social complexities, and that many deemed overweight by others are healthy, and have little desire or pressing need to lose weight.

The GLP-1 story has taken decades to play out, and at first fighting fat had nothing to do with it. In the early 1980s, researchers discovered GLP-1 while investigating diabetes and blood sugar regulation. Years of painstaking and sometimes discouraging work followed, but gradually the discoveries piled up, illuminating a hormone with expansive influence on the body and brain. Scientists learned that GLP-1 lowered blood sugar in people, and drug companies began to explore it as a diabetes treatment. In the 1990s, it emerged that injecting GLP-1 into the brains of rats made them eat less. A study of 20 healthy young men found that after a hearty breakfast, those getting intravenous GLP-1 infusions indulged less at a lunch buffet than those on a placebo.

The first GLP-1 drug was exenatide (Byetta), approved in 2005 for type 2 diabetes. Instead of the human hormone, its backbone was, improbably, a similar peptide in the venom of a giant lizard, the Gila monster. Almost 5 years later Novo Nordisk released liraglutide (Victoza), modeled on human GLP-1. It, too, was a diabetes drug, but in late 2014, the U.S. Food and Drug Administration blessed it for obesity.

The drugs didn't really catch fire until 2 years ago, when Novo Nordisk's next iteration, semaglutide, was greenlit for weight management in the U.S. (It's marketed as Ozempic for diabetes and Wegovy for obesity.) Unlike its forerunners, semaglutide required an injection just weekly rather than once or twice a day. And in a pivotal trial, people taking it lost an unprecedented 15% of their body weight over about 16 months. Many on the drug also describe a dampening of "food noise," the relentless and distressing desire to keep eating.

Since then, the frenzy has only intensified. According to electronic health records, 1.7% of people in the U.S. have been prescribed either Wegovy or Ozempic this year. (GLP-1 drugs are also approved in Europe for weight loss but availability varies.) Novo Nordisk's market value now exceeds the gross domestic product of Denmark, its home country. "When I look around this room I can't help but wonder: Is Ozempic right for me?" quipped comedian Jimmy Kimmel at the Academy Awards in March, poking fun at speculation over which movie stars took the drug.

But amid the jokes and soaring sales lurked a vital question. Could GLP-1 drugs actually safeguard health in people with obesity? This year brought an answer: yes.

In August, a trial of 529 people with obesity and heart failure found that after 1 year, people on semaglutide had almost double the heart improvement, as measured by a standard heart failure questionnaire, and could walk an extra 20 meters in 6 minutes compared with those in the placebo group. That same month, Novo Nordisk announced that in a much larger trial of 17,000 people with excess weight and cardiovascular disease, people on semaglutide had a 20% lower risk of fatal or nonfatal heart attacks and strokes than those on placebo; the study was published in November in *The New England Journal of Medicine*. The trials were the first to show in large numbers that GLP-1 drugs produced meaningful health benefits beyond weight loss itself. Meanwhile, a trial examining whether semaglutide delays kidney disease progression in diabetes patients showed such positive outcomes it was stopped early.

The reach of GLP-1 drugs is now widening in ways its inventors couldn't have imagined. Trials are underway for drug addiction, after people with obesity and diabetes described less longing for wine and cigarettes while on the treatment. Researchers theorize the drugs bind to receptors in the brain that mediate desire for other pleasures in addition to food. Clinical trials are also testing GLP-1 drugs to treat Alzheimer's and Parkinson's diseases, based in part on evidence they target brain inflammation.

But medical breakthroughs are rarely straightforward, and the ebullience surrounding GLP-1 agonists is tinged with uncertainty and even some foreboding. Like virtually all drugs, these blockbusters come with side effects and unknowns. Complications including nausea and other gastrointestinal problems lead some to abandon treatment. In September, U.S. regulators updated Ozempic's label to indicate a potential risk of intestinal obstruction, and in October, a Canadian team reported an increased chance of that complication as well as pancreatitis.

Doctors also worry about people who aren't overweight or obese turning to the treatment to slim down. A 2022 study reporting that semaglutide fueled 16% body weight loss in teenagers with obesity was met with hope but also hand wringing, as it underscored a vexing question: Are GLP-1 agonists "forever" drugs that people have to take indefinitely to preserve weight loss?

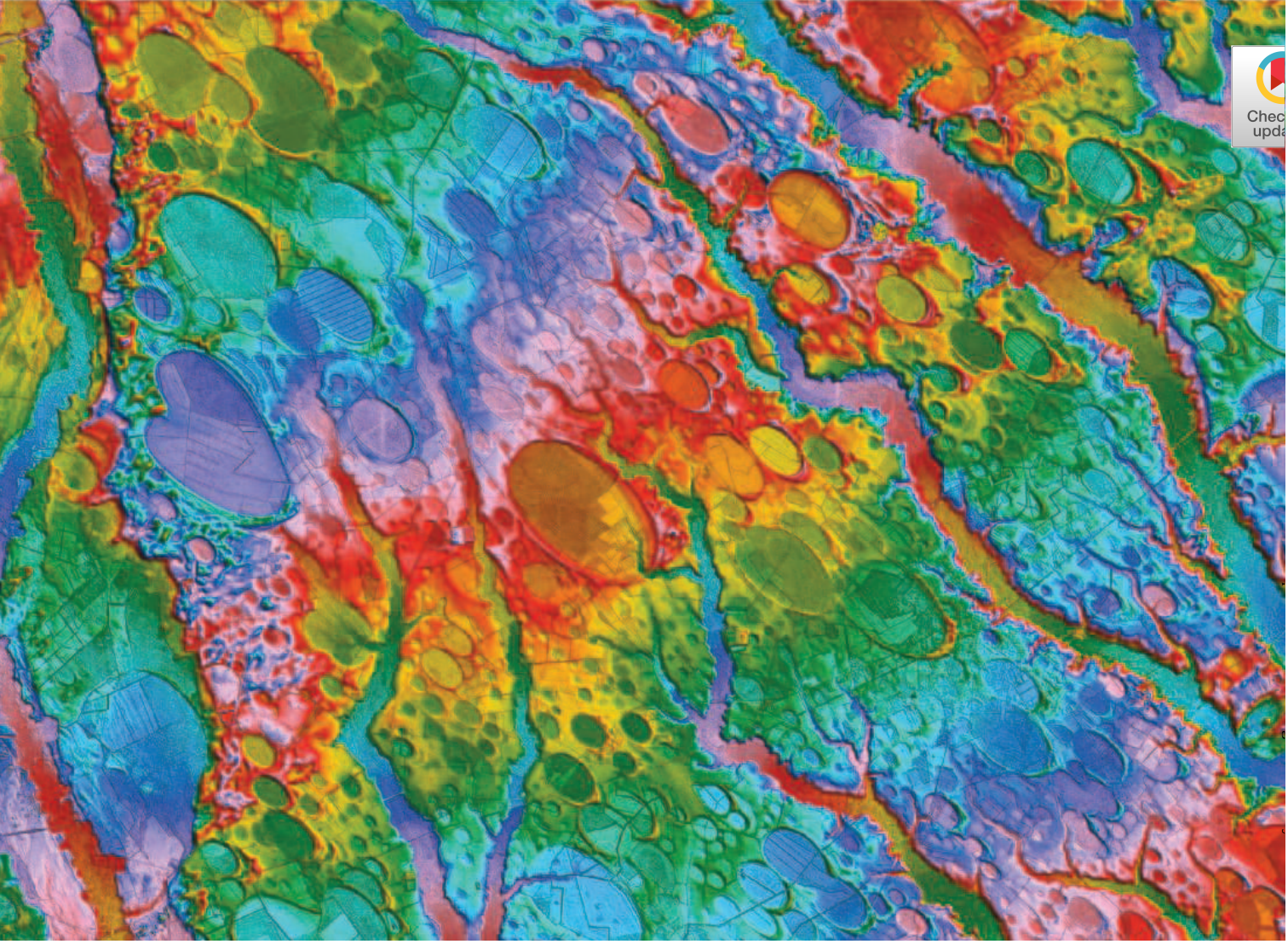
Right now it appears they may be, though the jury is still out. Researchers reported that 1 year after people stopped therapy, two-thirds of their lost body weight returned. For researchers who increasingly consider obesity a chronic condition, the need for ongoing treatment isn't surprising. But the drugs' cost can be prohibitive, with a sticker price of more than $1000 a month, and the prospect of lifelong use troubles many.

Against this backdrop, the next chapter is already unfolding: therapies that mimic multiple hormones and appear to be even more slimming. One, Eli Lilly & Co.'s tirzepatide, was approved in the U.S. in November for weight loss after being greenlit last year for diabetes; a large clinical trial reported that those taking it lost up to 21% of their body weight.

As the GLP-1 story continues, one thing is clear: These new therapies are reshaping not only how obesity is treated, but how it's understood—as a chronic illness with roots in biology, not a simple failure of willpower. And that may have as much impact as any drug. ■

VIDEO AND PODCAST
GLP-1 insights from the bench to the clinic at: https://www.science.org/boty2023

ILLUSTRATION: (OPPOSITE PAGE) STEPHAN SCHMITZ/FOLIOART

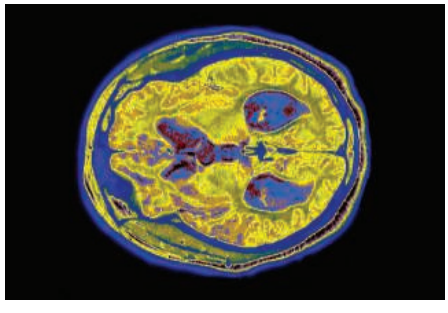


At last, modest headway against Alzheimer's

Medicine has had little to offer the tens of millions of people worldwide with Alzheimer's disease, and the few approved treatments have only targeted symptoms. But in January, U.S. regulators greenlit the first drug that clearly, if modestly, slows cognitive decline by tackling the disease's underlying biology; a second, related treatment is close behind. Neither comes close to a cure, and both have serious risks, but they offer new hope to patients and families.

The brains of people with Alzheimer's hold tangled protein clumps called beta amyloid, and for years scientists debated whether removing them would help patients. Various therapies that did so flopped. But the new treatment, an antiamyloid monoclonal antibody called lecanemab, slowed loss of cognition by 27%, compared with placebo, in a pivotal 18-month trial. It was enough to persuade regulators in the United States and later Japan to approve it. In trial results this summer, another antibody treatment that also targets brain amyloid, called donanemab, slowed cognitive decline by as much as 35% versus placebo in a slightly different patient population, and U.S. approval could come any day. Both therapies are given intravenously.

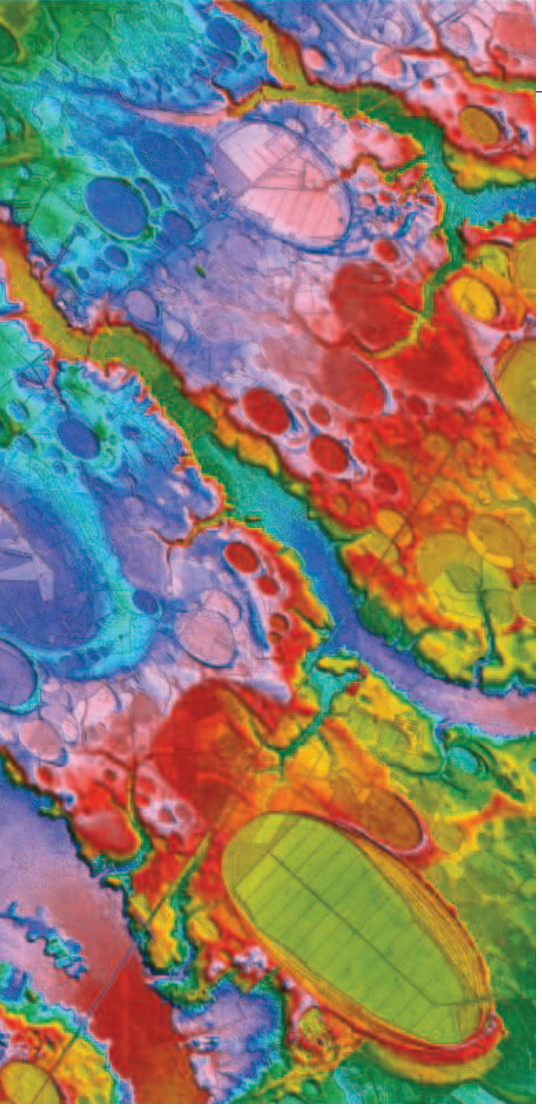
Although Alzheimer's researchers, doctors, and patients are celebrating, they're also eyeing a dark side: the risk of brain swelling and brain hemorrhage from the treatments, which in rare cases has been fatal. People with a gene variant predisposing to Alzheimer's, called *APOE4*, are especially prone to this side effect; also potentially at higher risk are people with Alzheimer's who take drugs to prevent or dissolve blood clots, including those who suffer a stroke and get powerful clot busters as emergency treatment.



New antibody therapies may slow neurodegeneration in the brains of people with Alzheimer's disease.

As the Alzheimer's community weighs the benefits and risks of antiamyloid drugs, it's also craving more data. One question is whether the modest improvement in cognitive slowdown grows with time on therapy. Another is whether these treatments, if given early enough to people at high risk of disease, can delay symptom onset.

The new therapies show that amyloid is a fruitful target, but they are just the beginning. In the coming years, scientists hope to figure out how to maximize their benefits—and find new treatments that work even better. —*Jennifer Couzin-Frankel*



Lidar maps of coastal North Carolina reveal kilometer-wide circular depressions that may encompass seeps of hydrogen.

Hunt for natural hydrogen heats up

In 1859, Edwin Drake sank 20 meters of cast-iron pipe into the earth beneath Titusville, Pennsylvania, and struck oil, collecting it in a bathtub. The well kicked off the U.S. oil rush and changed the world. This year saw the start of another energy rush, this one based on hydrogen produced naturally within Earth. Unlike oil the gas could be a tonic, not a toxin, for the climate.

Historians might one day trace its birthplace to another unlikely town: Bourakebougou, Mali. In 2012, engineers unplugged a borehole there that had been cemented shut in 1987, after a careless cigarette sparked an explosion. The gases it spewed turned out to be 98% hydrogen. A generator was hooked up. Producing only water as exhaust, it supplied the village with its first electricity. Curiously, after a decade of withdrawals, gas pressures in the borehole have not decreased—a suggestion that a deep source is replenishing the hydrogen.

Inspired by the discovery, prospectors are now finding signs of significant hydrogen deposits on every continent save Antarctica. Venture capital is flowing to startups such as Koloma, which came out of stealth mode in July with $91 million in funding, including investments from Bill Gates's Breakthrough Energy Ventures. In September, the U.S. Geological Survey (USGS) launched a research consortium with support from Chevron and BP, and the Advanced Research Projects Agency-Energy began a $20 million natural hydrogen R&D program.

That Earth holds any hydrogen at all defies conventional geological wisdom. Because hydrogen is energy rich and reactive, researchers thought that in Earth's crust most of it would be eaten up by microbes or converted into other compounds. Its surprising existence in so many places has prompted speculation that it leaks up from Earth's core or is created as radioactive elements in the crust split water. But many researchers believe it is generated when water reacts with iron-rich minerals at high temperatures and pressures.

An unpublished USGS study suggests Earth may hold 1 trillion tons of hydrogen—enough to satisfy growing demand for hydrogen as a fuel and fertilizer ingredient for thousands of years. Some prospectors say extracting it could prove far cheaper than manufacturing "green hydrogen" with renewable electricity, an approach supported by billions of dollars in government subsidies. But the big question is whether Earth's hydrogen is concentrated in reservoirs that companies can tap economically. If it is, environmentalists may find themselves in the odd position of cheering the roughnecks on with cries of "drill, baby, drill!" —*Eric Hand*

Early-career scientists rise up

For decades, graduate students and postdocs have complained about low pay and poor working conditions. Their frustration took center stage over the past year as early-career scientists banded together to demand changes in the system.

Last winter, 48,000 academic workers in the University of California (UC) system staged the largest academic strike in U.S. history, winning sizable pay increases for graduate students and postdocs. Such collective action has been especially pronounced in the United States, but in Canada, thousands of academic workers across the country engaged in a mass 1-day protest in May to demand increased federal funding for graduate students and postdocs. And in Germany, early-career researchers campaigned for changes to postdoc contracts.

ILLUSTRATION: DANIELVILLENEUVE/ISTOCK.COM

In the U.S., the unionization and strike activity has forced many universities to agree to increases in pay and other benefits, including child care allowances and improved workplace harassment policies. "We're doing this for us, but we're [also] doing this for the people that will come after us," says Álvaro Cuesta-Domínguez, a molecular biologist at Columbia University and an executive board member of a union that negotiated a new contract for the university's postdocs and associate researchers. "We need to provide better conditions for the future generation of scientists."

Early-career scientists have also put pressure on universities to enact changes by voting with their feet and leaving academia entirely. An increasing proportion are headed to more lucrative industry jobs after graduation, which has left many professors in recent years struggling to fill vacant postdoc positions.

Many faculty and university administrators agree changes are necessary. But it's been challenging to navigate the strain on their budgets. Professors, who often pay the salaries of early-career scientists out of their grants, may be forced to hire fewer graduate students and postdocs.

It remains to be seen whether funding agencies will increase grant support to pay for salary boosts for early-career scientists. In the meantime, some universities have moved to help faculty adjust to rising personnel costs. Researchers at UC Berkeley, for example, hope a combination of short-term measures to help professors pay for raises and longer term strategic planning and budgeting will ultimately lead to an ecosystem in which everyone can thrive. —*Katie Langin*

Footprints along an ancient lake in New Mexico may have been left 5000 years before archaeologists thought people arrived in the Americas.

Early peopling of the Americas steps closer to acceptance

The story of people in the Americas may have gained a new first chapter. In the prevailing picture, the first arrivals came from Asia via the land that once bridged the Bering Strait, then traveled down the Pacific Coast about 16,000 years ago. But this year, researchers came closer to confirming a remarkable claim that would push that date back at least 5000 years.

A handful of sites previously hinted that people may have made the journey earlier than the standard account supposed. Flaked stones and burned animal bones from southern Chile stretch back to 18,500 years ago, for example, and possible stone tools in a Mexican cave date to 26,000 years ago. But none offered unambiguous evidence of human activity, and most archaeologists remained skeptical.

In 2021, researchers working in White Sands National Park in New Mexico announced a potentially paradigm shifting discovery: unmistakable human footprints, left on the muddy shore of an ancient lake as early as 21,000 to 23,000 years ago. The team based those dates on seeds from a grassy aquatic plant that were found in layers surrounding the footprints and dated by radiocarbon. But there was room for doubt, because the seeds could have absorbed ancient carbon from sediments dissolved in the lake water, boosting their measured age. So the White Sands team redated the footprints using pollen from land plants and quartz grains embedded in sediments between and below the tracks. The new dates line up perfectly with the original paper, they reported in October.

If the dates are correct, the prints were left at the peak of the last ice age, when glaciers covered Canada, suggesting humans must have made the journey into the Americas before those ice sheets formed.

The new work has persuaded some initial skeptics. Others still wonder whether wind or erosion might have deposited older sediments on top of the footprints, making them appear more ancient than they are. But all are eager for more clues about the White Sands people, such as a hearth or stone tools, which could confirm their presence as well as provide hints about their culture. This year's redating could spark a re-evaluation of other contested sites and will likely send archaeologists racing to excavate other ice age sediments in search of confirmation—or even more surprises. —*Lizzie Wade*

Earth's carbon pump is slowing

If the world's oceans could be said to have a heart, it would reside in the Southern Ocean. There, in a few remote regions off the coast of Antarctica, salty surface waters plunge to the bottom of the ocean. The downgoing water pulls heat, oxygen, and carbon dioxide from the atmosphere and sequesters them in the abyss, where the water slowly spreads north. It is one of the primary components of the global overturning circulation, the massive flow that connects all the world's oceans, helping them capture one-third of humanity's annual carbon emissions. When it surges or falters, many researchers believe, this carbon pump can amplify climate change.

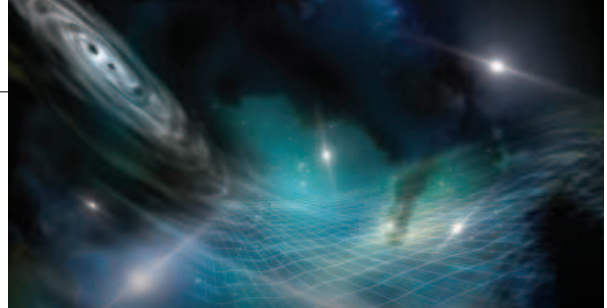
This year, several studies made clear the pump is in deep peril. The first troubling signs came years ago, notably from Deep Argo robotic probes, floats that autonomously roam down to 4000 meters. They found that the Antarctic bottom waters were warming and their volume was shrinking, both signs the current was slowing, allowing warmer waters above to intrude.

More direct evidence came from a study published in March in *Communications Earth & Environment* and led by U.S. researchers. They fed the sparse record of historical ship measurements in the region into a climate model, which showed the circulation has slowed by up to 20% since the 1970s. Then, in May, a study published in *Nature Climate Change*, led by Australian scientists, used measurements from ships and buoys to show that the flow of abyssal water slowed almost 30% from 1992 to 2017. Traditional climate models had forecast that the circulation might slow, but not for years. The new data indicate that a threat once thought to loom in the distance has already arrived.

There is little certainty so far on why this slowdown has occurred, how much humanity has driven it, or how it could affect climate. But the increased freshwater coming off Antarctica from melting ice sheets is likely to be a primary cause, making the surrounding waters fresher and less apt to sink. As global warming continues, the melting—and the slowdown—are sure to intensify. *—Paul Voosen*

Dense salty waters from the Weddell Sea help drive the Southern Ocean's deep circulation. Antarctic meltwater, however, appears to be slowing this current.

IMAGES: (LEFT TO RIGHT) A. TRAYLER-SMITH/GREENPEACE/PANOS/REDUX; AURORE SIMONNET/NANOGRAV COLLABORATION



The din of giant black hole mergers overheard

This year, astrophysicists picked up signs of a faint, long-sought cosmic rumble. It's the sound of gravitational waves from titanic masses in motion: supermassive black holes across the universe grinding around each other in tight orbital pairs. The observation—researchers are wary of calling it a "discovery" until they're certain—is both the strongest support so far for the existence of these monstrous duos and a powerful demonstration of a method that uses signals from distant stars to detect gravitational waves.

Supermassive black holes weighing millions or billions of times as much as our Sun reside at the heart of galaxies. When galaxies merge, their central black holes can end up locked in an ever-tightening orbit. The first stages of their death spirals are hidden from Earth-bound instruments. But when they approach within a few light-years of each other, their circular motion emits slow but powerful gravitational waves.

They can't be picked up by the Laser Interferometer Gravitational-Wave Observatory (LIGO), which in 2015 detected gravitational waves for the first time, from two star-size black holes merging (*Science*'s 2016 Breakthrough of the Year). To sense those millisecondslong waves, LIGO measured the distance traveled by laser beams down 4-kilometer-long vacuum tubes. But snaring the yearslong waves produced by supermassive black holes requires a far bigger net.

So astronomers turned to pulsars, burned-out stars that spin hundreds of times per second while firing out a jet of particles that emit radio waves. As that lighthouselike beam sweeps past Earth, radio telescopes register pulses as regular as an atomic clock. Over the past 2 decades, astronomers regularly monitored a few dozen of the most metronomic pulsars for tiny deviations from their rhythm. A passing gravitational wave will compress or stretch the space between the pulsar and Earth, subtly shifting when pulses arrive.

In June, five teams around the globe, each monitoring a different set of pulsars, jointly announced that with 15 years of observations they had beat down noise in the data enough to say that what remained was the combined din of supermassive black hole binaries across the universe, possibly millions of them. The teams are now hunting for yet more pulsars so they can map the hum and zoom in on galaxies where titanic black holes are slow dancing. *—Daniel Clery*

The AI weather forecaster arrives

For nearly as long as the modern computer has existed, it has been used to forecast the weather. First deployed during World War II to simulate nuclear weapons and artillery trajectories, nascent computers were soon adopted by meteorologists to simulate the future state of the atmosphere, creating the modern discipline of numerical weather prediction. And although that discipline has grown ever more sophisticated and now produces reliable forecasts several weeks in advance, its approach has remained the same: using massive amounts of computing power to solve the fluid dynamics equations governing the atmosphere.

Over the past year, artificial intelligence (AI) has begun to change that. Tech companies including Google, Huawei, and Nvidia have trained AI models to predict the weather up to 10 days in advance, with an accuracy rivaling or even topping traditional models—and with far less computational overhead. Rather than solving equations, these "deep learning" models predict the near future based on patterns learned through training on 40 years of past weather, as captured by observations fed through the numerical model of the European Centre for Medium-Range Weather Forecasts (ECMWF), the world's top forecaster. Once trained, the models can spit out a forecast on a desktop in 1 minute rather than taking 2 hours to run on a supercomputer. But as with most AI, no one truly knows what patterns they're learning.

ECMWF has already begun to produce its own AI forecast, and other weather agencies are scrambling to catch up. The new models aren't perfect. They struggle to predict certain essential features—hurricane intensity, for example. But researchers expect AI forecasters will only improve as they begin to learn from direct weather observations collected by sensors, not just data already passed through existing models. And their speed will likely allow forecasters to run them many times over, capturing the full spread of uncertainty that results from butterfly effects in the atmosphere.

No one expects traditional numerical weather prediction to disappear; climate models, for example, rely on the same equation-solving paradigm. AI may struggle to take over these forecasts because its models simulate changes in future that may not resemble the training data of the past. But in the long term, the output of supercomputer-driven climate models could itself become training data for a climate forecasting AI—which might ultimately outstrip its mentors.

—Paul Voosen

The maternal child ward at Kilifi County Hospital in Kenya sees many children with symptoms of malaria. A clinical trial of the new R21 vaccine took place in the city.

New hope against malaria

The effort to beat back malaria with vaccines got a double boost this year. A large evaluation of the world's first malaria vaccine, called Mosquirix, showed it significantly lowered deaths in young children, the group the disease hits hardest, killing nearly 470,000 a year in sub-Saharan Africa alone. Now, a second vaccine is poised to join the fight, with the approval by the World Health Organization (WHO) of a shot called R21/MatrixM. Similar to Mosquirix in design, it can be produced more cheaply and in greater quantities. It should help fill the huge gap between supply and demand for malaria vaccines, potentially preventing tens of thousands of children's deaths a year.

Mosquirix, also called RTS,S, has only modest efficacy, and its protection soon wanes. That's one reason WHO decided to roll out the vaccine under careful study. In 2019 it began a multiyear pilot phase in which nearly 2 million babies and young children in Ghana, Kenya, and Malawi received the vaccine. By 2021, preliminary results convinced officials it was safe and effective enough to approve for wider use. In October, WHO officials reported

Trained on 40 years of weather data and models, artificial intelligence can now forecast the paths of hurricanes with eerie accuracy.

comparisons between regions where the vaccine was offered during the pilot phase with control regions where it wasn't. The vaccine lowered the rate of severe malaria by 22%. Even more striking, mortality from all causes (except accidents) in children who were age eligible for the vaccine was 13% lower in areas where it was offered than in areas where it wasn't, demonstrating that even this imperfect vaccine can save lives. But GSK, which manufactures Mosquirix, can only produce 18 million doses between now and 2025, enough to fully vaccinate just 4.5 million children out of the 40 million born in malaria-affected regions each year.

R21 could help fill that gap. Developed at the University of Oxford, it is licensed to the Serum Institute of India, a major vaccine-maker. The company says it can produce 100 million doses per year at between $2 and $4 per dose, less than half the price of Mosquirix. Long-awaited data from a phase 3 trial of R21 involving 4800 children in four countries were released as a preprint in September. The results suggest that over the first 18 months the vaccine is at least as effective as RTS,S, and perhaps a bit more effective, though there have been no direct comparisons of the two vaccines. WHO says R21 should be available for widespread use by mid-2024. —*Gretchen Vogel*

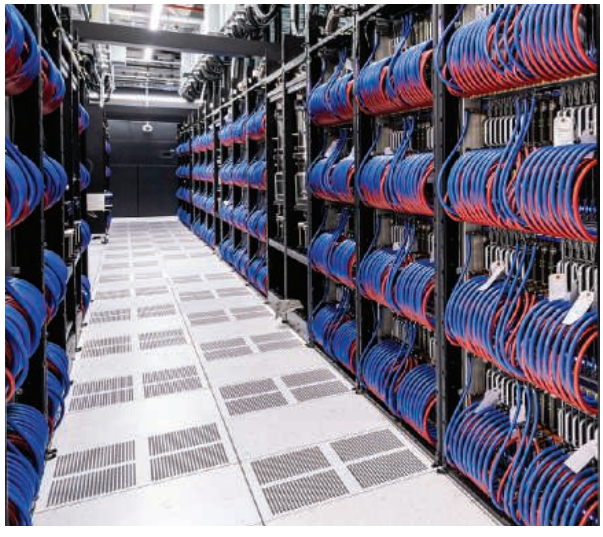
The dawn of exascale computing

More than a decade in the making, the era of exascale science finally arrived this year. Frontier, at Oak Ridge National Laboratory, became the first openly acknowledged exascale computer to open to scientific users, able to crunch through challenges in fields from climate to materials at an exaflops—1 quintillion (10^{18}) mathematical operations per second. (China is said to have had exascale machines in operation for years, but the country doesn't share details of its hardware.) Some 60 teams are expected to log time on Frontier next year. But early results have begun to pour out.

Among them: Frontier enabled researchers led by materials scientists at the University of Michigan to merge two theoretical frameworks for predicting the behavior of up to 600,000 electrons in a material with near-perfect precision, something previous efforts could only manage for about 1000 electrons. That allowed the team to simulate how defects in magnesium alloys form, grow, and move, an advance that should help spur the development of ultralight materials for fuel-efficient cars and airplanes.

Researchers at Lawrence Livermore and Sandia national laboratories also used Frontier to sharpen the resolution of the U.S. Department of Energy's primary global climate model. This, for the first time, made it possible to incorporate the physics that gives rise to clouds across the entire globe, an advance expected to greatly improve climate change forecasts. Last month, both those efforts were awarded the top prizes in their fields for supercomputing simulations.

The exascale explorations have only just begun. An exascale machine at Argonne National Laboratory is now undergoing final debugging before opening to users. Next year, new exascale supercomputers are expected to come online in California and Germany, with others in France and Japan set to follow—opening the doors to science on a previously unattainable scale. —*Robert F. Service*



Exascale computers like Frontier at Oak Ridge National Laboratory are bringing unprecedented computational power to many fields of science.

PHOTO: OAK RIDGE NATIONAL LABORATORY

Cuts in the U.S. Antarctic Program this year jeopardized projects such as one to monitor Weddell seals.

BREAKDOWNS OF THE YEAR

What went wrong in the world of science

U.S. Antarctic meltdown

Troubles mounted this year for the U.S. Antarctic Program (USAP), the global leader in supporting research on the frozen continent. The National Science Foundation (NSF), which manages the program, faced tough questions from lawmakers after a 2022 report documented "pervasive" sexual harassment. Some 72% of women who worked in Antarctica in recent years viewed sexual harassment as a problem, according to the report by an outside contractor.

Also this year, Antarctic scientists were stunned to learn that more than half of the USAP projects scheduled for the 2023–24 season had been canceled or reduced in scope. NSF officials place part of the blame on the pandemic, which has stretched out a $500 million renovation project at McMurdo Station, the program's hub on the continent, and reduced the number of available beds for scientists and those who provide essential logistical support. They also cite the rising cost of that support within a program budget that has remained flat.

Other looming challenges are the soaring cost of a new and more capable research icebreaker slated to replace an aging vessel and the expiration of a contract on a second, smaller vessel used for research cruises and to resupply a second U.S. station. Researchers say USAP's problems could cripple their ability to monitor a region that provides early warnings about the effect of climate change on the planet. They also fear that the diminished prospects for working in Antarctica could push early-career scientists into other fields with brighter futures. —*Jeffrey Mervis*

COVID-19 aftershocks

Continuing suspicions that the COVID-19 pandemic began with a leak from a government laboratory in Wuhan, China, fueled discord and distrust among scientists, intelligence analysts, and policymakers. Many scientists say compelling evidence supports the theory that the pandemic originated from a natural spillover from infected animals sold in a market in Wuhan. But China has hampered attempts to assess the spillover scenario—and not allowed an independent evaluation of the lab.

This year, non-Chinese scientists discovered publicly posted genetic sequences from environmental samples collected at the market by Chinese researchers early in the pandemic; they argue the sequences indicate infected animals were present at the market and support the spillover theory. The Chinese team that gathered the samples initially did not reveal the existence of the animals and then disputed that conclusion, as did those outside China who believe SARS-CoV-2 came from a Wuhan lab.

In the United States, an effort by nine federal intelligence entities to publicly explain their views on the pandemic's origins delivered little clarity; a majority favors

natural spillover, but with low confidence. In Congress, Republicans held fiery hearings that accused U.S. health officials of actively concealing a lab leak in order to cover up U.S. funding for bat coronavirus research at the Wuhan lab, but they offered no hard evidence. Meanwhile, fears of lab accidents have prompted U.S. proposals to more tightly regulate research with viruses that have pandemic potential, but some scientists fear new rules could hamper work on infectious diseases. —*David Malakoff*

Superconductor claims hit resistance

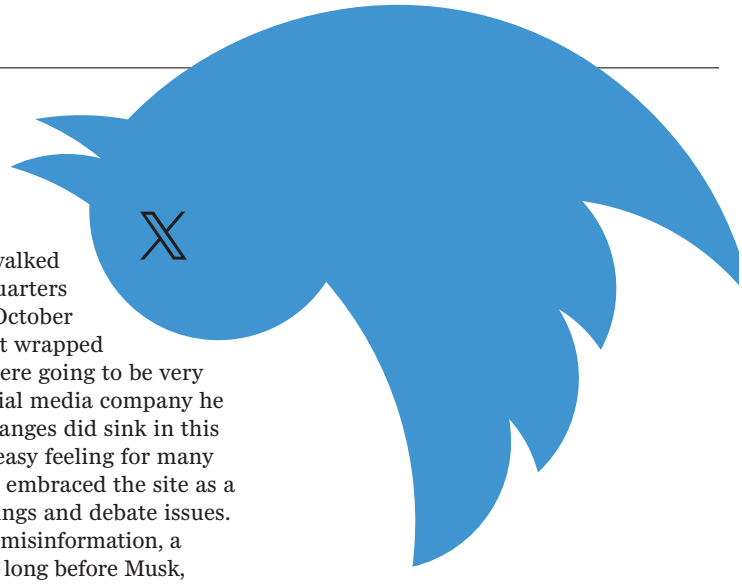
A claim that was hailed early in the year as "revolutionary" ended the year in tatters. In the 8 March issue of *Nature*, researchers led by Ranga Dias, a physicist at the University of Rochester, reported making a blue crystalline compound that when squeezed to moderate pressures turned red and exhibited a long-sought property: It conducted electricity without resistance at room temperature. But on 7 November the journal retracted the paper at the request of eight of the 11 coauthors, who said it did not accurately reflect their contributions and results. Dias did not say whether he agreed or disagreed with the retraction.

Superconductors are widely used in MRI scanners, specialized electronics, and particle accelerators. But most only work at ultracold temperatures, requiring expensive cooling equipment. Extreme pressure is one solution for some superconductors, but this, too, limits practicality. Room-temperature superconductors that work at or near ambient pressure could revolutionize electrical power transmission, save billions of dollars, and slash greenhouse gas emissions.

The November *Nature* retraction wasn't the only setback Dias suffered this year. On 7 August, *Physical Review Letters* retracted a 2021 Dias paper because of apparent data fabrication. That followed news in April that Dias may have plagiarized portions of his 2013 Ph.D. thesis at Washington State University. And all of this came after a *Nature* retraction last year of another near–room-temperature superconductivity paper by a Dias-led team. Dias has denied all allegations of misconduct.

Meanwhile, in August another claim of room-temperature superconductivity, from a South Korean team, went viral online, quickly followed by a backlash from skeptical scientists. Although no allegations of misconduct were made, the incident represents another high-profile misstep in the pursuit of the superconductor dream. —*Robert F. Service*

Research on Twitter goes down the drain

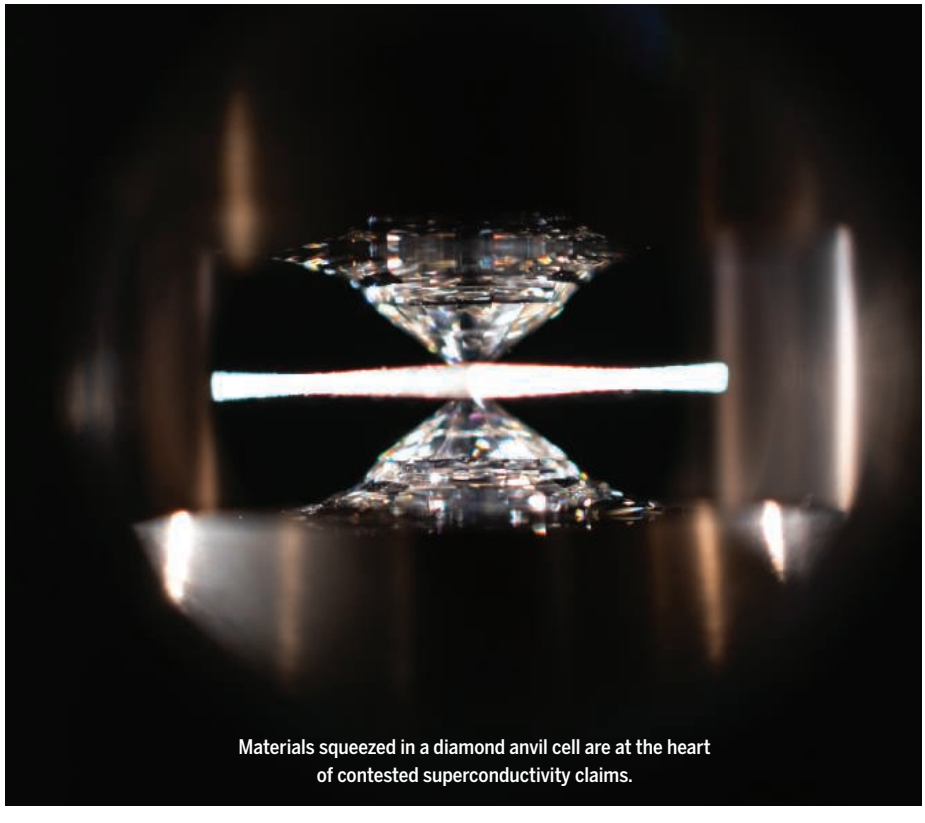


When Elon Musk walked into Twitter headquarters carrying a sink in October 2022 it was a threat wrapped in a joke: Things were going to be very different at the social media company he renamed X. The changes did sink in this year—as did an uneasy feeling for many scientists, who had embraced the site as a place to share findings and debate issues.

Hate speech and misinformation, a problem on Twitter long before Musk, increased even more as he scaled back content moderation, and many scientists left the site for good. A survey *Nature* conducted in July found that 7% of scientists using Twitter had stopped. Since then, Musk's own sharing of conspiracists' claims has persuaded many more researchers to follow suit. Some have set up digital camp at alternatives such as Mastodon, Bluesky, or Threads but others have not, and many online science communities have fractured.

More important, Twitter stopped sharing its data with researchers, shutting down its free application programming interface this summer and replacing it with one that gives more restricted access to data and is prohibitively expensive. Dozens of studies were suspended or pivoted to other platforms. The move hit particularly hard, because Twitter—historically more generous with its data than other social media platforms—had become a kind of model organism for those who study issues such as polarization and the spread of misinformation.

One silver lining for researchers: In the European Union, the Digital Services Act requires platforms like X to grant independent researchers some access to data beginning 1 January 2024. Some researchers have already applied, though how useful this avenue turns out to be will only become clear next year. —*Kai Kupferschmidt*



Materials squeezed in a diamond anvil cell are at the heart of contested superconductivity claims.

IMAGES: (TOP TO BOTTOM) X CORP.; ADAPTED BY C. SMITH/*SCIENCE*; MIKE BRADLEY/BIGSLIDE

EDITORIAL

More questions than answers

H. Holden Thorp
Editor-in-Chief,
Science journals.
hthorp@aaas.org

Weight loss and obesity have been challenging topics in science and public health for decades. While scientists have pursued understanding through studies of metabolism and the microbiome, public health scholars have shown that social factors such as poverty and inadequate health care are also major contributors. And though it is true that excess weight carries with it numerous comorbidities that lower life expectancy and add costs to the health care system, studies also show that it is possible to be overweight and healthy. That doesn't change the fact that obesity can be associated with poor mental health, in part because of the corrosive societal stigma that associates being overweight with mental weakness, not biochemistry.

Science's Breakthrough of the Year is the development of glucagon-like peptide–1 (GLP-1) agonists to treat obesity and weight gain, and this year's discovery that they can blunt obesity-associated health problems. GLP-1 agonists have been around since 2005 to treat type 2 diabetes and were approved for weight loss as early as 2014. But it was only since 2021, when the GLP-1 drug semaglutide was approved (as Wegovy) for obesity (it is sold as Ozempic for diabetes), that the frenzy really began, and mainstream media started referring to them as "the medical sensation of the decade."

As detailed in Jennifer Couzin-Frankel's news story, the results of clinical trials are impressive. In the pivotal trial, semaglutide was shown to lead to a loss of 15% of body weight in 16 months, and a later follow-up, tirzepatide, may work even better. In studies of cardiovascular disease and kidney disease in people with obesity or diabetes, GLP-1 agonists have shown potential benefits. At this point, it's safe to assume that there will be clinical trials for more indications in the coming years, including whether these treatments may ease addictions. The benefits may extend across a range of diseases, with some even resulting from primary effects of GLP-1 homeostasis and not just reduction of comorbidity due to weight loss.

> **"...for all their promise, GLP-1 agonists have raised more questions than they have answered—a hallmark of a true breakthrough."**

But for all their promise, GLP-1 agonists have raised more questions than they have answered—a hallmark of a true breakthrough. The first is about their cost and availability. These issues are likely to be worked out over the next few years, but for the moment, access to these drugs is limited. The cost can be over $1000 per month and is not always covered by insurance, if the medicine can even be obtained. Over time, the price may well come down as supply goes up, and insurance will likely cover the medication as more benefits are demonstrated that save on health care costs down the line—both in terms of catastrophic disease and the need for additional medications required to treat other conditions known to be associated with excess weight. The pressure is on—and should stay on—the drug companies and insurance providers to work this out.

Other questions relate to the need to stay on the drugs indefinitely to avoid regaining weight. A major assumption is that these are lifetime drugs. Are there safety implications associated with long-term use? How will clinicians decide how much weight loss justifies lifetime use of the drug? And at what age? These concerns become more pronounced for younger patients who will be on the drug for longer; at the same time, obesity among adolescents is a pressing problem on multiple fronts.

The development and implementation of these drugs are forcing important discussions about the way obesity is considered. Old pejorative tropes about obesity being the result of low willpower were hurtful to begin with, but now there is compelling evidence that a biochemical difference, not mental weakness, is responsible for weight gain. Reduced appetite and reduced "food noise" are clear benefits of these drugs. Scientists will over time refine our understanding of the genetic and environmental factors involved with body mass. That might lower the stigma and judgment around weight.

And that would truly be a breakthrough.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

10.1126/science.adn3693

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**

Aspen Pharmacare in South Africa is one of several companies hoping to help the continent produce most of its own vaccine doses.

HEALTH EQUITY

A billion-dollar boost for vaccinemaking in Africa

Gavi, the Vaccine Alliance, has committed up to $1 billion to bolster Africa's ability to sustainably produce its own doses of lifesaving vaccines. Manufacturers based in Africa produce only 1% of the vaccine doses used on the continent. Last week, Gavi announced that with money left over from the COVID-19 Vaccines Global Access Facility—an effort to provide an equitable distribution of COVID-19 vaccines—it would create the African Vaccine Manufacturing Accelerator (AVMA) to focus on preventing 11 priority infectious diseases. As an incentive to begin making the doses, AVMA will give a manufacturer $10 million to $25 million for each vaccine that the World Health Organization determines meets international quality standards. If Gavi and its partner, UNICEF, agree to buy the vaccine, AVMA will "top up" the price paid by as much as 50 cents extra per dose as an "accelerator payment." Its goal is to help at least four African manufacturers supply a total of more than 800 million doses over a decade.

EU agrees to landmark AI rules

REGULATION | After marathon negotiations, the European Union's three main governing bodies last week approved a comprehensive new law governing the development and use of artificial intelligence (AI) systems across the bloc. The world's first enacted legislation of its kind, the measure bans certain applications, such as the untargeted scraping of facial images from the internet and closed-circuit TV footage to create databases. It imposes the most restrictions on classes of AI categorized as posing a high risk of harm to health, safety, the environment, or fundamental rights. Companies that don't comply could be fined up to €35 million or 7% of global adjusted revenue. The regulation exempts systems used for defense, research, and innovation. To help smaller European companies stay relevant in the global race to develop AI, the law leaves it up to national authorities to set up "regulatory sandboxes" that would provide controlled environments for developing and training AI systems.

U.S. OKs gene editing for sickle cell

BIOMEDICINE | The U.S. Food and Drug Administration (FDA) has approved for the first time a medical treatment that uses the gene-editing tool CRISPR. The therapy offers a genetic fix for sickle cell disease, which is caused by defects in the oxygen-carrying protein hemoglobin. The product, Casgevy, developed by Vertex Pharmaceuticals and CRISPR Therapeutics, uses CRISPR to turn on a gene for the fetal form of hemoglobin in a lab-grown population of a patient's

harvested blood stem cells. The cells are then returned to the body. When FDA announced its decision last week, it also approved another sickle cell therapy—Lyfgenia, from maker bluebird bio—that adds a gene for adult hemoglobin to similar stem cells. The U.S. approval comes soon after the United Kingdom became the first country to approve a CRISPR treatment for sickle cell, on 15 November. For these potentially one-time treatments, Casgevy is priced at $2.2 million, Lyfgenia at $3.1 million.

HIV vaccine trial ends in failure

INFECTIOUS DISEASES | In an all too familiar disappointment, the only remaining HIV vaccine efficacy trial has been stopped early after poor results. Following failed trials of other HIV vaccines going back more than 2 decades, the PrEPVacc study began in 2020 and enrolled more than 1500 volunteers in South Africa, Tanzania, and Uganda. They received one of two combinations of up to four vaccines that present HIV proteins to the immune system in different ways. Study leaders hoped to report results next year. But an independent data monitoring committee notified them last month that results to date indicated no chance that either strategy being tested was protective. At an international AIDS conference in Zimbabwe on 4 December, PrEPVacc's leaders publicly announced the decision to stop vaccinating participants. As part of the trial, all of them also have received anti-HIV pills, known as pre-exposure prophylaxis (PrEP). The study will continue to provide PrEP and assess its impact on infection rates.

Macron names science advisers

POLICY | French President Emmanuel Macron last week established the country's first multidisciplinary scientific advisory council to help him "fully put science at the heart" of political decisions. (Other, specialized panels have advised the government on climate change and the pandemic.) Comprising 12 French scientists including two Nobel laureates and five women, the council will meet privately with Macron each quarter; its remit includes proposing new projects to build. Unlike some other countries, France does not have the position of chief scientific adviser.

Biden wants NIH to eye drug costs

TECHNOLOGY TRANSFER | President Joe Biden's administration last week proposed strengthening the leverage of U.S. federal agencies to force those who hold patents resulting from taxpayer-funded research to more broadly license them. Biden's move comes as he faces political pressure to help lower the high price of key medicines, such as the prostate cancer drug Xtandi. Under a 1980 law, owners of such patents may exclusively license them to companies; if public access is prevented, the government can force the patent owner to issue more licenses. Biden's proposal would clarify that the National Institutes of Health (NIH) can deem a drug's high price a barrier to public access. NIH has never exercised its "march-in rights," despite requests from advocates. Drug companies and academic institutions say it would dissuade industry from commercializing NIH-funded research. The Department of Commerce is accepting comments on the proposal until 6 February.

Fellows at the National Institutes of Health protested for better pay and working conditions outside the agency's main administration building in June.

PHOTO: MELISSA LYTTLE

LABOR RELATIONS

NIH workers vote to unionize

Early-career researchers at the U.S. National Institutes of Health (NIH) last week voted 98% in favor of forming a union—the first of its kind at a federal research agency, joining a growing wave of such unionization efforts at universities across the country. The bargaining unit, called NIH Fellows United, is set to represent more than 5000 nonpermanent researchers, including graduate students, postdocs, and research fellows, who work at NIH research facilities. Under federal rules, workers won't be permitted to engage in strike activity. But organizers hope the union will push for better pay and working conditions.

AAAS strips harasser of honor

#METOO | AAAS said this week it has revoked the honorary fellowship of molecular biologist David Gilbert after his former institution found he engaged in severe, pervasive sexual harassment. The society's 8900 active fellows are recognized for distinguished achievements and normally elected for life. In 2018, AAAS began to allow removal of the title for documented scientific misconduct, including harassment. *Science*, which AAAS publishes, reported in April that Gilbert's former employer, Florida State University, made the determination of harassment in 2020; the next year, he moved to the San Diego Biomedical Research Institute, which Florida State did not inform about its finding. In San Diego, one trainee reported being harassed, but a probe by the institution found no "unlawful harassment."

Australia buys water for ecosystem

CONSERVATION | Australia's government on 7 December enacted a law allowing it to buy more water from farmers to improve the ecology of the country's largest river system, the Murray-Darling Basin. Scientists have documented that low water flows have caused chronic harm, including algal blooms and fish kills. The law guarantees the basin a quantity of water promised under a 2012 policy. Under a previous policy, the government capped the amount of water rights it bought, which hindered the ecosystem's recovery.

HIGHER EDUCATION

Florida law blocks hiring of Chinese students

Top officials would have to approve job offers to research staff—and process remains opaque

By **Jeffrey Mervis**

A new state law is thwarting faculty at Florida's public universities who want to hire Chinese graduate students and postdocs to work in their labs.

In effect since July, the law prohibits institutions from taking money from or partnering with entities in China and six other "countries of concern." The list of banned interactions includes offering anyone living in one of those countries a contract to do research. Students could be hired only if they are granted a waiver from the state's top higher education body. But how that process would work is not clear, and the 12 public colleges and universities covered by the law are still writing rules to implement the statute.

More than 280 faculty members at the University of Florida, which has the state's largest research portfolio, have signed a petition urging UF to clear up the confusion and to voice support for an open-door policy on hiring. "We urgently request a timely decision that allows us to recruit top international graduate students with an assistantship, irrespective of their nationality," declares the petition, sent on 6 December to UF President Ben Sasse and senior UF leadership. "Failure to act swiftly may result in the loss of exceptional students to other universities, and the damage will be irreversible."

The uncertainty created by the law has already kept faculty from making offers to research students in China for the fall of 2024, which would normally happen in December and early January. "We have missed that window," says Zhong-Ren Peng, a UF professor of urban planning who leads a center for adaptation planning and design. "And the best students cannot wait. Instead, they will go somewhere else."

Passed unanimously by the state legislation in May, the law (SB 846) is part of a broader effort by state lawmakers and Governor Ron DeSantis (R) to thwart what they see as attempts by China and other countries to influence what happens on campuses and steal intellectual property. Backers point to a few cases in which Florida scientists failed to disclose ties to China or used public funds to support work there as evidence that state universities need to be more vigilant.

The law applies to all academic interactions with China, Russia, Cuba, Iran, North Korea, Syria, and Venezuela. Exceptions are allowed only when the Board of Governors, which oversees higher education in the state, decides that the interaction "is not detrimental to the safety or security of the U.S. or its residents."

Such a waiver would be granted on a case-by-case basis for those seeking a research position as a graduate assistant or postdoc. (Students from the countries of concern could still be admitted to a graduate program but could not be paid to work on a research project.) A 2021 law already requires Florida universities to review the proposed employment of any foreign researcher, a category that covers not just people currently living in another country, but also anyone with 1 year or more of training or employ-

Corrected 14 December 2023. See full text.

In May, Florida Governor Ron DeSantis (R) signed bills restricting interactions with China.

hold U.S. citizenship. What's more, those turned away because of something in their background seen as suspect must be reported to FBI or another law enforcement agency designated by the governor.

In October, the Board of Governors provided guidance intended to help universities draw up policies that comply with the new law. The new procedures were supposed to be in place by 1 December. But universities are still scrambling to write them. "The university is presently developing policies and procedures that conform to these new requirements," says David Norton, UF's vice president for research.

Norton says UF now employs about 350 graduate assistants and 200 postdocs from the seven countries of concern. But UF faculty see a much bigger ripple effect from the new law. "A substantial number of skilled applications for our graduate programs originate from these countries of concern," the petition notes, singling out China and Iran. "Restrictions on recruiting could significantly reduce our applicant pool … and lead to a significant erosion of UF's standing within those international communities."

Computational chemist Wenjun Xie, who took a tenure-track position at UF in July, worries that eliminating students from countries on the list could cripple his ability to staff his lab and generate the type of preliminary results needed to win a federal research grant. "As a new investigator working in the rapidly evolving field of AI research, this situation is particularly problematic for me," he says.

Chenglong Li, a chaired professor in the same department of medicinal chemistry that recently hired Xie, fears the restrictions "will hurt UF's reputation at a time when we are striving for excellence in artificial intelligence." Li adds, "I think we are shooting ourselves in the foot."

Li also worries the law could put faculty supported by federal grants at odds with U.S. laws banning discrimination by those receiving federal funding. "As a principal investigator, we promise not to discriminate based on race or national origin," he notes. "But SB846 is doing exactly that."

Singling out individuals is never a good idea, agrees Rebecca Keiser, who leads a new office on research security at the National Science Foundation. "There are risks posed by a particular entity or a particular government," says Keiser, who declined to comment on the Florida law. "But I think it's better to focus on how to get to yes, with the proper safeguards in place, rather than on excluding individuals." ■

FACILITIES

Their budget bursting, U.S. particle physicists dream small

New effort to study big bang afterglow heads to-do list

By **Adrian Cho**

Particle physicists in the United States have released a plan for the next decade that looks less like a child's wish list and more like a parent's cautious budget. Although some physicists dream of exotic new colliders, the report of the ad hoc Particle Physics Project Prioritization Panel (P5) lists just five, mostly smaller projects. That's because the U.S. program, supported by the Department of Energy (DOE), is still busy with a massive neutrino project that has greatly exceeded its initial cost estimate and is behind schedule.

Still, other physicists are encouraged. "This is better than I expected," says Daniel Akerib, a particle physicist at SLAC National Accelerator Laboratory. "Even given the constraints, they found a way to fit new things in."

The product of more than a year of deliberation, the report, presented on 7 December to DOE's standing High Energy Physics Advisory Panel (HEPAP), represents the consensus of the panel's 31 particle physicists, says chair Hitoshi Murayama, a theorist at the University of California, Berkeley. "We never voted on anything."

The first recommendation sets the tone, says Regina Rameika, associate director for DOE's high energy physics program, which has a $1.17 billion budget this year. The highest priority, the report says, is to "complete construction of projects and support operations of ongoing experiments." In other words, Rameika says, "We've got to finish what we've started."

Those commitments include various experiments at Fermi National Accelerator Laboratory (Fermilab), massive underground detectors that are striving to detect hypothetical particles of dark matter called weakly interacting massive particles (WIMPs), and a 4-meter telescope to probe the mysterious dark energy that appears to be accelerating the expansion of the universe. They also include a planned upgrade to increase the intensity of the proton beams at the world's biggest atom smasher, Europe's Large Hadron Collider (LHC), to which DOE will contribute parts worth $650 million.

Above all, physicists must complete a huge two-piece experiment, the Long Baseline Neutrino Facility (LBNF) and the Deep Underground Neutrino Experiment (DUNE). It would fire neutrinos from Fermilab in Illinois to a massive subterranean detector 1300 kilometers away in South Dakota. Neutrinos, nearly massless and almost undetectable, come in three types and can morph from one to another as they zip along at light speed. LBNF/DUNE aims to provide the ultimate study of such neutrino oscillations, which might shed light on how the nascent universe generated more matter than antimatter.

In 2014, researchers expected LBNF/DUNE to cost less than $1.9 billion and to

PHOTO: RYAN POSTEL/FERMI NATIONAL ACCELERATOR LABORATORY

Excavation in South Dakota for the huge Deep Underground Neutrino Experiment is nearly done.

be completed in 2028. The cost has grown to $3.2 billion—and that's not for the whole project. Original plans called for the detector in South Dakota to include four modules, each containing 17,5000 tons of frigid liquid argon. In March 2022, DOE officials split the project into phases, with the first consisting of just two detector modules. The others would be added later, as an upgrade. Neutrinos won't flow until 2031.

Nevertheless, P5 recommends starting five projects costing more than $250 million apiece. The top priority would be an ensemble of new telescopes at the South Pole and in Chile's Atacama desert called CMB-S4 to study the afterglow of the big bang, the so-called cosmic microwave background (CMB). DOE and the National Science Foundation would split the $800 million cost of the project, which would scrutinize the CMB for evidence that the newborn universe underwent an exponential growth spurt called inflation.

The panel says the U.S. should also go ahead with LBNF/DUNE phase two and prepare to participate in an international "Higgs factory"—an electron-positron collider that would produce a particle called the Higgs boson in copious amounts—should researchers in Europe or Japan decide to build it. And it endorses an ultimate WIMP detector weighing up to 80 tons and an expansion of the existing IceCube detector sunk into the South Pole ice, which detects ultra–high energy neutrinos from space.

To make all this affordable, P5 is banking on some luck, both technical and political. On the technical side, Fermilab researchers have realized that relatively simple changes to the lab's proton accelerators might increase by 75% the number of neutrinos generated for LBNF/DUNE. That implies the second phase might do the same amount of science with one additional module for the distant detector, not two, cutting costs.

The P5 report also assumes that by 2027 the budget for DOE's high energy physics program will grow by 34%—the boost Congress promised in the 2022 CHIPS and Science Act—and will increase with inflation thereafter. If that boost does not materialize, plans will have to be scaled back, it says.

P5 also calls for aggressive R&D to develop ideas such as a collider that would smash muons, heavier, fleeting cousins of the electron. In principle, a muon collider could reach higher energies than a Higgs factory or the LHC while remaining much smaller.

The idea is "compelling," says Heidi Schellman, a physicist at Oregon State University and a member of HEPAP. But she cautions that because muons decay so rapidly, such a collider presents myriad challenges. "Somebody needs to 'splain to people how hard it is to make a muon collider." ■

ASTROPHYSICS

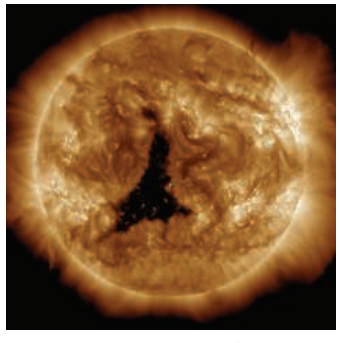
Do tiny black holes from cosmic dawn hide within giant stars?

Idea could account for universe's mysterious dark matter

By **Adam Mann**

Grunge music: a source of validation for a generation of disaffected youth—and a surprising source of scientific inspiration for Earl Bellinger of the Max Planck Institute for Astrophysics. While listening to Soundgarden's 1994 hit *Black Hole Sun* 2 years ago, he contemplated a curious question: Might itty-bitty black holes from the dawn of time be lurking in the hearts of giant stars?

The idea is not so far-fetched, he found. Astronomers could detect these trapped black holes by the vibrations they cause on the star's surface, and if numerous enough, they could function as the mysterious dark matter that knits the universe together, Bellinger and colleagues find in a study published this week in *The Astrophysical Journal*. "It is speculative, but interesting," says Juan García-Bellido, a theoretical physicist at the Autonomous University of Madrid.



A black hole buried in the Sun might affect vibrations at its churning surface.

Ordinary black holes are born when large stars die and their cores collapse into objects so dense that even light can't escape their gravity. In 1971, famed physicist Stephen Hawking proposed another origin. In the thick soup of particles present moments after the big bang, certain spots might have been dense enough to collapse into black holes ranging in size from the microscopic to the incredibly huge.

If pervasive enough, these primordial black holes could function as the dark matter thought to make up 85% of the matter in the universe. Astronomers have searched for them by looking for flashes that would arise when they eclipse a distant object and magnify its light like a lens. None have been spotted. But if a black hole was tiny enough, with a mass of an asteroid and a diameter as small as a hydrogen atom, the flashes would be too dim to be picked up in such surveys.

Bellinger and his colleagues considered a universe in which dark matter was made entirely of such teensy black holes. They found some would get trapped within the gas clouds that give birth to stars, and would end up sinking to the cores of newly formed stars.

At first, very little would happen. Even a dense stellar core is mostly empty space. The most microscopic of the black holes would have a hard time finding matter to consume, Bellinger says. "It could take longer than the lifetime of the universe to eat the star."

But larger ones, about as massive as the asteroid Ceres, could grow in a few hundred million years. Material would spiral onto the black hole, forming a disk inside the star that would heat up through friction and glow. Once the black hole was as massive as Earth, it would churn up the core like pot of boiling water, cause the star's outer layers to puff up and redden, and become the main source of brightness. "It will become a black hole–powered object rather than fusion-powered object," says study coauthor Matt Caplan, a theoretical physicist at Illinois State University. The team calls these entities "Hawking stars."

To identify them, Bellinger says, astronomers could tune in to the particular frequencies at which the stars vibrate. Because a Hawking star would churn throughout its interior, rather than just in the topmost layers like an ordinary red giant, it would be expected to thrum with a particular combination of frequencies. Such waves could be detected in the way the star's light pulses and throbs. Bellinger is applying for funding to study a list of about 500 promising nearby red giants to see whether any shudder with the signs of a black hole.

And if he were to one day find a star that seems to harbor a primordial black hole? "That would be great," he says. "I joked with some people that it would be the dumbest ever Nobel Prize, to discover dark matter because you're inspired by *Black Hole Sun*." ■

Adam Mann is a science journalist in Oakland, California.

ILLUSTRATION: ROBERT NEUBECKER

RESEARCH ETHICS

Universities uneasy with plan to disclose misconduct findings

Transparency could benefit, but some see a threat to privacy

By **Jeffrey Mervis**

The U.S. government wants to strip away some of the secrecy surrounding university investigations of research misconduct by publicizing findings that institutions might otherwise keep confidential. Advocates of greater transparency say the change, proposed by the federal Office of Research Integrity (ORI), would boost public confidence in research and correct the scientific record sooner.

"Good for [ORI]," says lawyer C.K. Gunsalus, who leads a center on research ethics at the University of Illinois Urbana-Champaign. "The government provided the money to do the research," she says. "And if the institution that received the money concludes that the work wasn't up to the rigor expected, the public has a right to know."

But many university administrators dislike the proposed change, announced in October and now open for public comment. Because the notice doesn't spell out what information ORI might disclose or when it would use its new authority, they worry the information released might violate privacy laws or distort the actual findings.

"ORI isn't taking into account state, local, or institutional privacy or confidentiality considerations," says Kris West of the Council on Governmental Relations, which monitors federal research policy for its hundreds of institutional members. She and others also question the legal basis for having ORI publicize the findings of another, outside organization. Such authority, says attorney Minal Caron of Ropes & Gray, who counsels universities conducting misconduct investigations, "would be inconsistent with the practices of other federal agencies."

Created in 1992, ORI monitors the conduct of scientists funded by the National Institutes of Health and other agencies within the Department of Health and Human Services. It defines research misconduct as fabrication, falsification, or plagiarism.

Currently, institutions must notify ORI if they find misconduct and prescribe corrective action, such as retracting a paper. ORI can then begin a process leading to its own finding of misconduct, which it makes public—or it can let the matter rest, which often means a finding is never revealed. The proposal would give ORI a third option: Simply announce what an institution has done.

Eugenie Reich, a Boston attorney who represents whistleblower scientists suing institutions over alleged research misconduct, thinks the new authority "will be an incentive for universities to make their findings public." She says ORI "is basically telling them that if you don't, we might."

The proposed language on disclosure is part of the first overhaul of rules ORI adopted in 2005. Expected to be finalized next summer, the changes would give universities less time to decide whether to pursue an allegation against a faculty member, require more record-keeping, and prohibit institutions from quickly closing a case they believe reflects "honest error." ORI's director, Sheila Garrity, has touted the changes as a "better way" to do business, and the office has received favorable comments along with questions about how some changes would be implemented.

The disclosure provision is drawing particularly strong reactions. ORI's notice says it would disclose institutional findings only "to protect the health and safety of the public … or to conserve public funds." It promises not to include "the names or other identifying information of the respondent(s)," referring to the scientist found guilty of misconduct.

Susan Garfinkel, vice president for research compliance at Ohio State University and former head of ORI's investigative unit, sees some value in disclosure. "People want to know if the paper was flawed, and if some of the data in it can't be trusted," she says. "So I'd support announcing that the image in figure six was fabricated."

But she and many other research administrators worry about a loss of control over what is an institutional product. They also fret that the information released could mischaracterize the overall findings or provide enough clues about identities to damage the reputations of lab associates or co-authors not involved in the misconduct. "These reports are hard to write, and university committees struggle to get the wording just right," Caron says. "So a brief summary could be very misleading, and leave a permanent dark cloud hanging over the head of someone who's not at fault."

Garfinkel and West say citing the problematic paper would make preserving anonymity impossible. That makes Garfinkel uncomfortable about disclosure when ORI decides not to issue its own finding of misconduct. With university-only findings, she says, "Sometimes you want to give people a chance to recover from it, and stay in science. But that's hard to do once you've been named."

Garrity declined to provide more details about the policy. But she says she welcomes feedback from interested parties, who have until 4 January 2024 to submit comments.

Garfinkel hopes her colleagues will offer enough pushback to convince ORI to remove the language before issuing the final rules. But Gunsalus thinks critics are overstating the potential risk of ORI's new authority. "Some universities have released their findings, and the sky hasn't fallen," she says. "Sunshine is the best disinfectant. So let us see the work." ■

BIOMEDICINE

Proposed FDA rule on lab tests sparks debate

Scientists divided over whether increased regulation would protect patients or limit access

By **Jennifer Couzin-Frankel**

The U.S. Food and Drug Administration (FDA) wants to tighten regulation of certain lab tests, and its proposal is touching off anxiety. Tests designed and used for patient care within a single lab have so far been largely exempt from FDA's review process. In September, however, the agency proposed a new rule that could subject roughly half of such tests, which number at least in the tens of thousands and range from prenatal screening to gene sequencing of tumors, to scrutiny. Comments from doctors, professional groups, research scientists, and others flooded in—almost 7000 by the time the scheduled comment period closed last week.

What exactly FDA will require of testmakers is uncertain. But the prospect of change is exposing schisms over whether more regulation will safeguard patients from inaccurate results or discourage innovation and reduce access to tests. "There's this layer of hysteria," says John Pfeifer, a surgical pathologist at Washington University School of Medicine in St. Louis. "You hear a lot of people saying this will put a lot of labs out of business." Pfeifer is one of many physicians at his university who routinely runs lab-developed tests on patient samples, and he supports additional regulation.

FDA already vets in vitro diagnostic tests that are developed by and used across multiple labs, including continuous glucose monitors and pregnancy tests run in doctor's offices. Decades ago, the agency opted not to scrutinize lab-developed tests, because they were mostly simple, small-scale assays. But in recent years, highly complex tests, such as sophisticated analyses of DNA, have galloped ahead. Their reach now often extends far beyond one hospital's patients; some are marketed by companies and offered nationwide. And the tests increasingly come with high stakes, including diagnosis of serious diseases such as cancer and heart disease.

FDA has long suggested these tests need vetting, too. It now proposes that they be regulated as medical devices, which are categorized as low-, medium-, or high-risk. The middle-risk category typically requires oversight from FDA prior to marketing, and the agency reviews the highest risk candidates for safety and effectiveness. FDA estimates that about half of lab-developed tests would fall into one of these two categories, and between 6% and 7% would need that more intense vetting.

Many researchers support some increase in oversight. The testing field has seen a "gradual slide" into commercialization and marketing of tests with sometimes unproven—or unpublished—accuracy, and a risk of potential harm, says Michael Mina, a clinical pathologist and epidemiologist, and chief science officer of eMed, which provides in-home tests and links patients to treatment.

But many, including Mina, fault the proposal as impractical and overly burdensome. FDA is "biting off more than they can chew" given the number of tests it would have to vet, Mina says. In February 2020, when Mina was at Harvard University, he and his colleagues developed a COVID-19 polymerase chain reaction test but were prevented from launching it until FDA reviewed evidence of its efficacy—even as one of the first superspreader events unfolded at a Biogen conference in Boston. A monthlong backlog of applications for "essentially identical tests" hindered public access to them, he says. (FDA notes in the new proposal that some of the COVID-19 tests submitted to it were unreliable.)

Another common concern is that intensive clinical testing before approval isn't always feasible. At Brigham Young University, epigeneticist Tim Jenkins is developing a test to predict Alzheimer's disease risk years before symptoms emerge based on detection of DNA from certain brain cells in a person's blood. A company he co-founded, Resonant, aims to launch the test early next year. "We have people emailing us every day" asking for this test, Jenkins says. He and his colleagues published a paper in October in *Frontiers in Neurology* validating the test in 50 people, and more validation is ongoing. But, he says, "If I put on my perfect study design cap," generating ironclad evidence could take a decade or more—dooming the test.

Pfeifer, however, worries about quality control issues. His team tested whether 19 academic and commercial labs could identify 54 gene variants in a colon cancer tissue sample. Patients with any of the variants don't respond to a particular cancer drug.

As the group reported in 2021 in the *American Journal of Clinical Pathology*, seven labs aced the test, detecting every variant correctly. Three made fewer than five mistakes, and nine had five or more errors—mostly false negatives, meaning a patient would be given a drug from which they would not benefit and might suffer side effects. The labs were privately notified of the results. Data like these, he says, "suggest that the current practices or procedures in place aren't adequate."

FDA proposes a phase-in of the new rules over 4 years, but its specific demands aren't yet clear. The draft document is peppered with questions on which the agency sought input, including whether a different policy should apply to labs in academic medical centers, as opposed to those in other settings, like commercial labs.

Whether the proposal will ever become a final rule is also uncertain. FDA tried a similar regulatory change in 2014, only to abandon it with the shift to a new presidential administration. Marsha Michie, a bioethicist at Case Western Reserve University who has studied lab-developed tests in prenatal care, hopes FDA goes through with the overhaul this time. "There's no question that regulation imposes additional burden on laboratories," she says, but "the fact that … it takes a little more time and effort to get these things to market is not necessarily a bad thing." ■

Robert DePalma examines asteroid ejecta at the fossil site, which records the hours after impact.

SCIENTIFIC COMMUNITY

Dinosaur extinction researcher guilty of research misconduct

But Robert DePalma did not commit fraud in paper claiming asteroid hit in springtime, university report finds

By **Michael Price**

Paleontologist Robert DePalma, whose high-profile claim that the asteroid that killed the dinosaurs struck in springtime drew accusations of fraud, is guilty of "several counts of poor research practice" that "constitute research misconduct," according to an investigation. The report by the University of Manchester, which it shared with *Science*, says DePalma did not fabricate data, but notes he was unable to say where the key isotope data underlying his 2021 paper in *Scientific Reports* were produced.

"I'm happy the university acknowledged poor research practices [and] misconduct," says paleontologist and Ph.D. student Melanie During. "That's a big win." She and her adviser, paleontologist Per Ahlberg of Uppsala University, had filed the fraud complaint, and she continues to think *Scientific Reports* should retract the paper by DePalma's team. "Nonscientific information has been published as scientific information, and it is still out there," she says.

DePalma emphasizes that the university, where he is now a Ph.D. student, cleared him of fabricating data. He says what he calls errors in "record-keeping and data presentation … collectively were labeled as a form of misconduct of a negligent sense, which is fuel for self-improvement and positive adjustment."

A representative for Springer Nature Group, which publishes *Scientific Reports*, told *Science* they were "considering our next steps." The paper bears an editor's note saying, "The reliability of data presented in this manuscript is currently in question."

The paper's conclusion—that the asteroid hit in springtime 66 million years ago—is not at issue. The finding comes from Tanis, a site in North Dakota leased by DePalma that preserves a trove of fossils evidently deposited on the day of the catastrophe 66 million years ago. During, who worked alongside DePalma at Tanis in 2017, published the springtime finding with Ahlberg and colleagues in *Nature* in February 2022, 2 months after DePalma's paper. In allegations to *Science* last year and in a preprint, During and Ahlberg alleged that DePalma had fabricated data in order to publish the claim first.

To determine season of death, both During's and DePalma's teams reported isotopic signatures taken from fossilized fin bones of paddlefish thought to have been killed minutes to hours after the asteroid's impact. In modern paddlefish bones, seasonal dietary changes lead to higher levels of carbon-13 during spring. During and Ahlberg found similarly high levels in the ancient fish.

DePalma's graphs of isotope levels also show the springtime spike, but the paper does not include the raw isotopic data. In their preprint and complaint, During and Ahlberg say irregularities in the graphs suggest data fabrication.

The university's Research Governance, Ethics and Integrity office investigated. Both During and Ahlberg appealed the results, which haven't been released. The appeal panel's report, released to *Science*, concludes DePalma drew the points in the graphs by hand from "an interim data sheet" given him by his now-deceased co-author Curtis McKinney. The irregularities "were explained as genuine errors resulting from the lack of raw data as a consequence of the death of McKinney," the report says.

McKinney worked at Miami Dade College, which lacks the type of mass spectrometer needed. "Although McKinney's institution did not have the kind of apparatus supposedly used for the analysis, this was not evidence that McKinney had not sent the samples elsewhere for analysis," the report says. It also cites "evidence that the isotope data existed in 2017, before During visited Tanis, so there appeared to be no motive for fabricating the data."

The report faults De Palma for "poor research practice" and for not being transparent about "the fact that he did not know in what laboratory [the isotope analysis] had been conducted, how and by whom." Together, these failings "constitute research misconduct," the university writes.

The panel absolved two Manchester researchers who co-authored the 2021 *Scientific Reports* paper: paleontologist Phillip Manning and geochemist Roy Wogelius. It notes that they were not supervising DePalma while he prepared the isotope graphs and did not supervise the isotope work.

Other isotope experts say they'd like to know where the data came from. "In my experience, it is highly irregular … [to] not include any primary data tables and instead rely solely on hand-traced figures (particularly with no disclosure of that choice)," writes Thomas Cullen, an isotope researcher at Auburn University, in an email.

John Eiler, an isotope analysis expert at the California Institute of Technology, says the data are "not only inconsistently reported, but also implausible." The report notes that an outside geochemist told the appeals panel the presented isotope data were "too good to be true." Eiler adds that "the inconsistencies re: McKinney's involvement were never addressed. … I would not accept this issue as settled."

The report concludes by saying the university will give DePalma the opportunity to redo the experiments. ■

PHOTO: BBC

GENOMICS

The history of the Coast Salish "woolly dogs"

DNA and traditional knowledge reveal the history of an extinct dog bred for its wool

By **Ludovic Orlando**

In many societies, dogs represent more than mere companions—they are highly esteemed working partners and integral community members that are sometimes revered for their spiritual importance (*1*). Dogs played these multifaceted roles in the Indigenous communities located around the Salish Sea, a region that encompasses tens of thousands of square miles in the Pacific Northwest near present-day Vancouver, Canada, and Seattle, Washington. However, the Coast Salish peoples also developed something rather distinct—a tradition of wool weaving based on the particularly thick and woolly coat of a special lineage of dogs. On page 1303 of this issue, Lin *et al.* explore the genome sequence of a Coast Salish dog named Mutton who died in 1859. This analysis was combined with traditional knowledge provided by local Indigenous communities to glean information on the breeding practices and genetic modifications that underlie this exceptional hairy dog phenotype (*2*).

Mutton accompanied the ethnographer George Gibbs as he participated in the Northwest Boundary Survey (1857 to 1862), which sought to solve border issues between the United States and Britain. Mutton's pelt was later preserved in the collections of the Smithsonian Institution in Washington, DC (see the photo). Lin *et al.* used samples of the pelt to obtain a genome sequence in which each nucleotide was read slightly more than three times on average. The authors compared the genome sequence against those from other animals, including a nonwoolly dog that died in the region (Semiahmoo Bay in the Pacific Northwest) in 1858, multiple dogs that lived ~5000 to 500 years ago before the arrival of European colonizers [precontact dogs (PCDs)] (*3*), and modern coyotes and dog breeds distributed worldwide.

Mutton's mitochondrial DNA (mtDNA) was similar to that of the other PCD specimens and seemed to be most similar to the mtDNA of a dog that lived in the Pacific Northwest ~1500 years ago. This indicates a direct local maternal line of descent because mtDNA is inherited from mother to offspring. By contrast, the nonwoolly dog from the Salish Sea region carried another mtDNA haplotype that originated from Europe, which largely replaced haplotypes of PCD lineages as the Americas were colonized. Nuclear genome data confirmed that Mutton's genome was mainly derived from the genomes of PCDs but also had ~16% European ancestry. This indicated a recent history of admixture in his genealogy, proba-

The pelt of Mutton, a Coast Salish wool dog who died in 1859, is preserved in the collections of the Smithsonian Institution in Washington, DC.

bly within the 18 to 47 years before his birth.

The persistence of ~84% PCD ancestry in Mutton's genome contrasts with the ubiquity of European ancestry that was present at that time in dogs brought over by colonizers and among the dogs that now populate the Americas (*3*). It indicates that Coast Salish peoples successfully maintained the genetic integrity of their dog lineage for a substantial period after first contact with colonizers. This, and the relatively limited differences identified between the two copies of each of Mutton's chromosomes, echoes traditional knowledge collected as part of the study that indicates that the woolly dogs were bred in isolated pens on coastal islands to reduce the risk of cross-breeding with other dogs.

Lin *et al.* also identified genes in which Mutton accumulated an excess of nonsynonymous mutations, which affect the encoded protein and, thus, potentially have biological consequences. They filtered out genes showing nonsynonymous mutations in modern dog breeds and other PCDs to obtain a candidate list of the genetic changes that are responsible for the hair phenotype of the Coast Salish dogs. A literature survey revealed that 28 of the candidate genes were involved in hair follicle function and hair growth. These include KN motif and ankyrin repeat domains 2 (*KANK2*); the homologous human gene is involved in a rare congenital abnormality of the scalp that causes a "woolly hair" phenotype (*4*). However, the mutations that are associated with hair modifications in modern dog breeds were largely absent from Mutton's genome. Therefore, the Coast Salish communities crafted a lineage of Indigenous dogs with distinct biological characteristics. The precise timeline of this development is uncertain, but archaeological and genetic evidence suggest origins of the woolly dog dating back at least ~2000 to 5000 years.

The study from Lin *et al.* is exemplary for its engagement with local Indigenous communities. Traditional knowledge keepers were not just consulted for legal approval—a common concern for many studies—but their insights were recorded throughout the process as research material and directly embedded in the main text of the article. By doing this, the authors follow in the footsteps of another study published earlier this year in which a group of Western and Indigenous scientists worked together to confirm First Nations of the American Southwest as earlier users of horses than classically assumed (*5*). Both studies demonstrate the power of engaging with Indigenous communities and the enriched research experience it offers, from design to data interpretation and dissemination.

In cross-cultural collaborations, mutual respect and equal footing are crucial factors, and it is essential to acknowledge and address potential challenges. For instance, the sharing of traditional knowledge comes with responsibility and must follow traditional protocols to ensure Indigenous sovereignty and avoid the misuse of information once disclosed. Indigenous concepts without an exact English translation may require nuanced explanations that are not always compatible with the word limitations of scientific publishing. Companion articles published alongside the original scientific article may offer better exposure for these explanations than supplementary materials (*6*). Furthermore, the publication structure and formats established by Western science may not align with traditional modes of oral transmission; direct quotes may help overcome this and ensure that Indigenous collaborators can speak in their own voices.

Indigenous science embraces a considerably diverse array of dimensions, or realms. Therefore, there is no one-size-fits-all solution when it comes to developing successful cross-cultural collaborations. However, studies like that of Lin *et al.* can provide inspiring templates for scientists who are increasingly concerned about the risks of scientific colonialism and the adoption and/or replication of research practices that undermine the contributions of their Indigenous partners (*7*). By acknowledging traditional knowledge as the product of a multigenerational peer-reviewing process within communities, these studies effectively share the scientific practice across the whole team, preventing an unbalanced power structure. Applying these ethics to the study of animals, rather than just human ancestors, promotes a more holistic framework to extend the focus of discussions for an ethical future of ancient DNA research (*8*–*10*).

Dogs were first introduced into the Americas ~15,000 years ago (*3*), but European imperialism, which started in the Americas with the first voyage of Christopher Columbus in 1492, ultimately resulted in the extinction of the Coast Salish dogs. When the last of Mutton's relatives lived remains unclear, but one study detected dog keratin peptides in the fibers of Coast Salish blankets that were made in 1862 (*11*). Future work should focus on understanding the collapse of the Coast Salish dogs and the broader extinction of PCDs as European dogs spread alongside their colonizing masters across the Americas. European colonizers also brought an array of other domestic species that originated on a different continent, with massive ecological and cultural repercussions that warrant attention. Research that views Indigenous communities as both agentive and resilient actors and investigates the way that they adapted their lifestyle in the face of oppressive and marginalizing policies as well as widespread epidemics may offer models for dealing with upcoming environmental crises. ■

Centre for Anthropobiology and Genomics of Toulouse, CNRS UMR5288, Université Paul Sabatier, Toulouse, France. Email: ludovic.orlando@univ-tlse3.fr

REFERENCES AND NOTES

1. P. Cunningham-Smith, K. Emery, *J. Ethnobiol.* **40**, 409 (2020).
2. A. Lin *et al.*, *Science* **382**, 1303 (2023).
3. M. Ní Leathlobhair *et al.*, *Science* **361**, 81 (2018).
4. Y. Ramot *et al.*, *J. Med. Genet.* **51**, 388 (2014).
5. W. T. T. Taylor *et al.*, *Science* **379**, 1316 (2023).
6. M. Hunska Tašunke Icu *et al.*, "Standing for *Unči Maka* (Grandmother Earth) and all life: An introduction to Lakota traditional sciences, principles and protocols and the birth of a new era of scientific collaboration" (Maxwell Museum Technical Series no. 42, Maxwell Museum of Anthropology, 2023).
7. M. C. Ávila-Arcos *et al.*, *Front. Genet.* **13**, 880170 (2022).
8. S. Alpaslan-Roodenberg *et al.*, *Nature* **599**, 41 (2021).
9. R. E. Fleskes *et al.*, *Annu. Rev. Genomics Hum. Genet.* **23**, 627 (2022).
10. E. Kowal *et al.*, *HGG Adv.* **4**, 100161 (2023).
11. C. Solazzo *et al.*, *Antiquity* **85**, 1418 (2011).

10.1126/science.adm6959

GLOBAL WARMING

Coral reefs in peril in a record-breaking year

Climate change and its impacts on coral reefs have reached unchartered territory

By **Ove Hoegh-Guldberg[1], William Skirving[2,3], Sophie G. Dove[1], Blake L. Spady[2,3,4], Andrew Norrie[2,3], Erick F. Geiger[3,5,6], Gang Liu[3], Jacqueline L. De La Cour[3,5,6], Derek P. Manzello[3]**

The upper ocean is undergoing unprecedented changes in conditions, ecosystems, and communities. These changes can be traced back to the early 1980s, when mass coral bleaching first appeared. Marine heatwave (MHW)–driven events are strongly correlated with rising sea surface temperature (SST) and climate cycles such as El Niño–Southern Oscillation (ENSO). SSTs in 2023 have been starkly different because extreme MHWs have engulfed much of the eastern tropical Pacific (ETP) and wider Caribbean. Many Caribbean reef areas experienced historically high heat stress that started much earlier (1 to 2 months) and was sustained for longer than the usual recorded seasonal changes. Patterns of SST from the past 40 years indicate that unprecedented mass coral bleaching and mortality will likely occur across the Indo-Pacific throughout 2024. These trends will worsen unless greenhouse gas (GHG) emissions decrease, with coral-dominated ecosystems likely to face substantial losses, leading to long-term damage to ecosystems and people across Earth's tropical regions (*1*).

Earth experienced its warmest days since 1910 in early July 2023, as well as the warmest month for SSTs (see the figure and fig. S1). Anomalously high SSTs in the ETP and wider Caribbean are more extensive now than in any other year on record (which started in the early 1980s). Individual coral reefs reached record levels of heat stress up to 12 weeks ahead of previously recorded peaks. Heat stress puts immense pressure on vital but fragile tropical ecosystems such as coral reefs, mangrove forests, and seagrass meadows. For example, Newfound Harbor Key, a coral reef in the Florida Keys, accumulated heat stress that was almost three times the previous record, with this occurring 6 weeks ahead of previous peaks (see the figure and fig. S2). This trend has occurred at all reefs in the Florida Keys and at many reefs throughout the wider Caribbean.

Historical data suggest that the present MHWs that are occurring throughout the ETP and wider Caribbean will likely be the precursor to a global mass coral bleaching and mortality event over the next 12 to 24 months as the El Niño phase of ENSO continues. ENSO is a recurring climate pattern that involves periodic changes in the temperature of waters in the central and eastern tropical Pacific Ocean, where El Niño is a warm phase that causes higher SSTs and La Niña is the cool phase. The difference between each ENSO phase (a natural cycle) is being amplified by climate change (*2*). The present El Niño conditions are predicted to persist for at least the first quarter of 2024. Since 1950, an El Niño state that is present for the last 6 months of a given year and for at least 3 months in the following year has occurred seven times. In every instance except one, the second year was the warmer of the two; the cooling effect of the Mount Pinatubo eruption in the Philippines during June 1991 thwarted this pattern in 1991–1992. Since 1997, every instance of these El Niño pairs led to a global mass coral bleaching and mortality event. The global extent of heat stress is indicated by degree heating week (DHW). DHW is a satellite-based metric of heat stress that measures the accumulation of daily SST anomalies over a 12-week period, with values at or above 4 indicating coral bleaching (see the figure). Crucially, 2023 is the first year of a potential pair of El Niño years with the warmest average global SST for February to July on record.

Warming seas and associated MHWs disrupt the intricate balance of marine ecosystems, such as coral reefs, that affects the distribution, health, and abundance of many species and ecological processes. Mass coral bleaching occurs when corals become stressed from factors such as SST, which causes them to lose their brown microbial symbiont, turning them white. At low-stress levels, the symbionts may return to coral tissues over a few months. Increasingly, stress levels are killing corals irrespective of their bleaching status. Seabirds and marine mammals that require specific temperatures for breeding, feeding, and migration are also at risk, as are large-scale biogeochemical cycles, such as the nitrogen cycle that drives nutrient generation (*1*, *3*, *4*). Furthermore, the increased prevalence of harmful algal blooms and the spread of invasive species and diseases are additional consequences of warming seas, which can have cascading effects throughout the food web and associated marine ecosystems. Outbreaks of coral disease have occurred for months to years after all prior mass bleaching events in the wider Caribbean, which has often led to higher rates of mortality than the initial heat stress (*5*). These effects, coupled with the present record-breaking MHWs in the ETP and wider Caribbean, are a poignant reminder of the risks of unchecked GHG emissions and attendant global warming (*3*, *6*).

Until now, efforts to curtail GHG emissions to the Paris Agreement objectives (i.e., maintaining global warming well below 2°C and aligning with 1.5°C over the long term) are failing. Instead, the global temperature trajectory is trending toward a 3°C increase over preindustrial levels by 2100 (*2*). This spells dire consequences for coral reefs and many other ecosystems, given the ongoing repercussions at 1.1°C of warming and the anticipated impacts that are likely at 3°C (*2*, *6*).

The persistence of coral reefs beyond the next few decades remains in serious jeopardy (*3*). Rising sea temperatures, coupled with other stressors such as ocean acidification and pollution, have severely weakened the resilience of coral reefs. Changes are insidious yet fundamental. Limited reproductive success of corals has been documented in the wider Caribbean for more than 20 years (*7*), making these reefs more vulnerable to disease outbreaks and extreme weather events. For example, a single thermal event in the Galápagos Islands in 1982–1983 drove levels of coral mortality to 95 to 99% and the complete loss of coral reef structures over 10 years with no recovery (*8*). Since May 2023, mass coral bleaching has occurred in multiple countries in the ETP (Mexico, Costa Rica, El Salvador, Colombia, and Panama) and in much of the Caribbean and Florida, which illustrates that a large-scale heat-stress event is underway.

Higher bleaching tolerance within coral species or across reef zones is associated with generally modest increases in thermal tolerance [i.e., +1°C in the Florida Keys (*9*)] or has

[1]School of the Environment, University of Queensland, St Lucia, QLD, Australia. [2]ReefSense, Townsville, QLD, Australia. [3]Coral Reef Watch, Satellite Oceanography and Climatology Division, Center for Satellite Applications and Research, US National Oceanic and Atmospheric Administration (NOAA), College Park, MD, USA. [4]College of Science and Engineering, James Cook University, Townsville, QLD, Australia. [5]Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD, USA. [6]Cooperative Institute for Satellite Earth System Studies, University of Maryland, College Park, MD, USA. Email: oveh@uq.edu.au

occurred in locations that have not yet experienced the severe heat stress that is expected over the next few decades [e.g., Panama (*10*)]. Furthermore, increased bleaching tolerance in corals because of an association with heat-tolerant algal symbionts (*Durusdinium* spp.) often comes with substantial physiological costs, such as reduced calcification and carbon transfer (*11*).

Some initiatives that focus on increasing bleaching tolerance through the introduction of heat-tolerance genes into natural populations ("assisted evolution") have shown some promise (*12*). However, scaling up such restoration efforts remains challenging, especially when bleaching resistance is not necessarily connected to coral survival (*4*, *13*). The rapidly increasing rate of SST rise exacerbates challenges because corals struggle to keep up with increasingly hostile conditions. As the intensity of MHWs has escalated, some historically dominant reef-building coral groups (e.g., staghorn *Acropora* species) have been observed to die abruptly (i.e., they slough tissue) rather than die from the loss of symbiotic dinoflagellates (through bleaching).

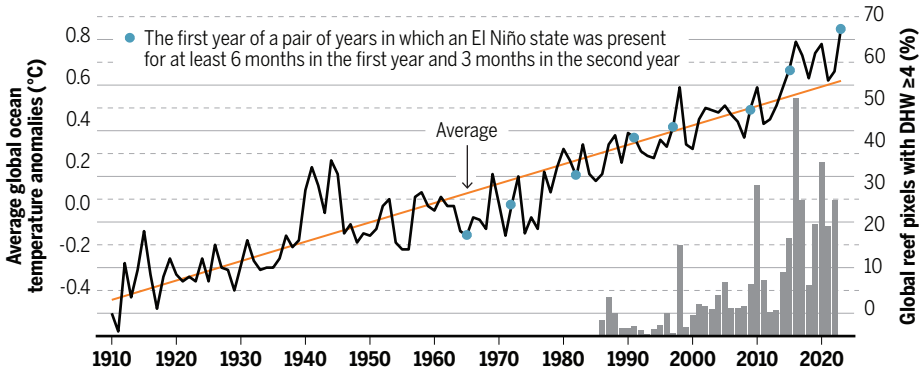
Consequently, identifying reef stress through observations of mass coral bleaching (especially with aerial surveys) will become increasingly challenging as corals decrease and specific coral groups die without bleaching. Management metrics must also move from bleaching-dominated measurements to those that focus on the death of corals and downstream consequences. Better mortality assessments require improvements in spectral analysis to separate skeletal epilithic algal growth from living coral tissue. Additionally, extension of the geographical scale and frequency of in-water reef surveys, by using advancements in robotics and image analysis through artificial intelligence, are needed. These improved in-water observations will also provide an important resource for improving satellite monitoring products. Stress monitoring metrics, such as DHW, would benefit from improved spatial and temporal resolution (e.g., to 100-m resolution at subdaily timescales). Large-scale coral mortality is much more destructive to reefs than large-scale bleaching. Dead coral skeletons are broken down by the combined effects of physical, biological, and chemical erosion, with recovery impeded by the instability of the resulting reef framework. As the architecturally complex, three-dimensional reef structures are eroded, the vital habitat that many coral reef–associated species depend on disappears, leading to ecosystem collapse (*4*). Importantly, this could undermine the persistence of as much as 25% of ocean biodiversity (*2*).

New approaches and strategies are urgently needed to mitigate the impacts of climate change, which have been recognized as a major threat to coral reefs for more than 40 years (*1*). Given the magnitude and speed of the changes to oceanic conditions, a comprehensive system should integrate measures to reduce GHG emissions and limit further warming and acidification of the oceans. Additionally, efforts to enhance the resilience of marine ecosystems, including protecting critical habitats and refugia, are equally important. Actions must integrate marine- and land-based approaches, including establishing and enforcing regulations within protected areas, promoting sustainable fishing practices, and reducing coastal pollution (*2*, *14*). Adaptive management practices are also crucial, given rapid environmental changes. Monitoring programs that track oceanic conditions, ecological indicators, and species distributions can provide valuable insights for decision-making and conservation planning. Furthermore, fostering international collaboration and knowledge exchange is essential for sharing best practices, lessons learned, and technological innovations that can aid in preserving marine ecosystems.

Identifying and safeguarding coral reef sites that are least exposed to climate change offers a scalable and targeted response that can assist in prioritizing investments and guiding adaptive management. Climate exposure varies between locations owing to local oceanography and other factors (*15*). Reefs that are the least exposed to ocean warming can be prioritized and protected from stressors unrelated to climate change, such as pol-

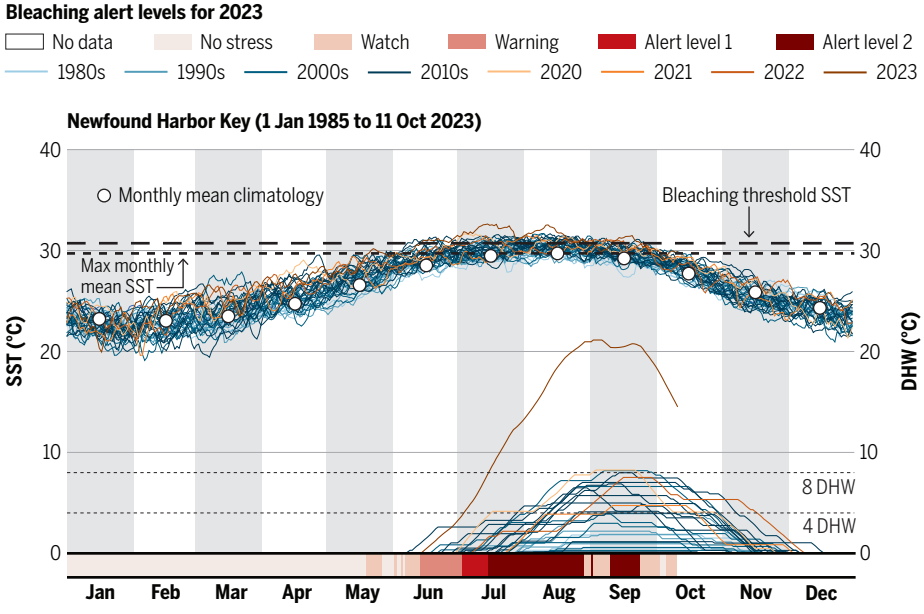
Coral heat stress

Average global ocean temperature anomalies from February to July each year (1910 to 2023) are increasing. This is coupled with increased periods of degree heating weeks (DHW) ≥4 (when coral bleaching is likely to occur). Data are from the National Oceanic and Atmospheric Administration (NOAA; www.ncei.noaa.gov), using climatology between 1901 and 2000 as the reference.



Newfound Harbor Key

At Newfound Harbor Key, annual sea surface temperature (SST) and DHW show substantial increases between 1985 and 2023. Data are from the NOAA's Coral Reef Watch version 3.1 (https://coralreefwatch.noaa.gov).



GRAPHIC: D. AM-PHAN/*SCIENCE*

lution, overfishing, and unsustainable coastal development. Once climate conditions stabilize, less-exposed reefs could serve as a source of reproductive materials (larvae), aiding corals in reclaiming these invaluable tropical seascapes (*15*). Such strategizing assists in addressing the scaling challenge, particularly with the limited resources available to manage the vast expanses of coral reefs that are being affected globally.

Given the complex and interconnected nature of marine ecosystems such as coral reefs, a comprehensive approach is necessary for mitigating the impacts of changing oceanic conditions. This approach should include rapid and scalable efforts to reduce GHG emissions as a global emergency, thereby limiting further warming and acidification of the oceans. Simultaneously, exploring technologies for preserving coral stocks and other important reef organisms for future dissemination is critical. It is essential to remain open-minded to new ideas while acknowledging the challenges inherent in addressing the global issues facing coral reefs at the appropriate scales and costs. Nonetheless, swift action needs to be taken to safeguard 25% of ocean biodiversity and the future of more than half a billion people worldwide who depend on coral reefs. ■

REFERENCES AND NOTES

1. O. Hoegh-Guldberg, *Mar. Freshw. Res.* **50**, 839 (1999).
2. Intergovernmental Panel on Climate Change (IPCC), "Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change," Core Writing Team, H. Lee, J. Romero, Eds. (IPCC, 2023).
3. O. Hoegh-Guldberg *et al.*, *Science* **365**, eaaw6974 (2019).
4. S. G. Dove *et al.*, *Commun. Earth Environ.* **1**, 55 (2020).
5. J. Miller *et al.*, *Coral Reefs* **28**, 925 (2009).
6. IPCC, *Climate Change 2022: Impacts, Adaptation and Vulnerability*, H. O. Pörtner *et al.*, Eds. (Cambridge Univ. Press, 2023).
7. T. P. Hughes, J. E. Tanner, *Ecology* **81**, 2250 (2000).
8. P. W. Glynn *et al.*, *Mar. Pollut. Bull.* **133**, 717 (2018).
9. D. P. Manzello *et al.*, *Glob. Change Biol.* **25**, 1016 (2019).
10. A. M. Palacio-Castro *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2202388120 (2023).
11. D. Tye Pettay *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7513 (2015).
12. M. J. H. van Oppen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2307 (2015).
13. T. P. Hughes *et al.*, *Nature* **556**, 492 (2018).
14. J. M. Gove *et al.*, *Nature* **621**, 536 (2023).
15. H. L. Beyer *et al.*, *Conserv. Lett.* **11**, e12587 (2018).

ACKNOWLEDGMENTS

The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect the views of the NOAA or the Department of Commerce. E.F.G., J.L.D.L.C., W.S., and B.L.S. were supported by NOAA grant NA19NES4320002 (Cooperative Institute for Satellite Earth System Studies) at the University of Maryland–ESSIC. O.H.-G. and S.G.D. were supported by the University of Queensland, Australia. Funding for Coral Reef Watch is provided by NOAA's Coral Reef Conservation Program.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk4532

Published online 7 December 2023
10.1126/science.adk4532

ASTRONOMY

The drivers of massive star evolution

An elusive population of stripped binary stars is finally revealed

By **Jon Sundqvist** *and* **Hugues Sana**

Although the Sun appears lonely in space, most stars born with higher masses are locked into cosmic dances with at least one companion (*1*). This pairing, however, comes neither without risk nor without consequence. Theoretical models predict that for many stars with birth masses 8 to 25 times that of the Sun, their outer layers are stolen by nearby partners (*2*). Yet despite the high predicted numbers, detection of such stripped stars has been elusive. On page 1287 of this issue, Drout *et al.* (*3*) report electromagnetic observations that finally reveal the missing population, confirming a key prediction of binary star evolution. Additional analysis of the stars' physical properties by Götberg *et al.* (*4*) further supports the idea that these systems are progenitors to neutron star binaries, including those recently detected by their bursts of gravitational waves.

Stars come to life within contracting gas clouds in interstellar space. Most are born with masses around or lower than that of the Sun. Occasionally though, the stellar nursery produces much more massive and luminous objects. Unlike the Sun, which will shine for another 5 billion years, these cosmic beasts live fast and die young, exhausting their nuclear energy source in just a few million years (*5*). Across the Universe, these rare massive stars serve as beacons of light, powering galaxies such as the Milky Way and allowing astronomers to observe stars at cosmological distances (*6*). A quantitative understanding of how massive stars live and die is thus important well beyond stellar physics, influencing research fields as diverse as the evolution of galaxies and the early cosmos, the origin and frequency of gravitational wave bursts, and the production of chemical elements essential for life.

In the classical picture of massive star evolution, a star is born in isolation and lives most of its life fusing hydrogen within its deep core. From the outside, it looks bright and blue because of its hot surface temperature of several 10,000 K. When core-hydrogen becomes depleted, the star expands, and its outer layers cool off. Depending on the exact mass and detailed physics (much of it still insufficiently understood) of the stellar interior, surface, and gas outflow, the evolved object may then directly explode as a supernova or shrink and heat up again before meeting its demise (*3*, *5*). Either way, the massive star ultimately ends up as a tiny and dense neutron star or a black hole, alone in a dark cosmos.

Over the past years, however, evidence has accumulated that this basic scenario may need considerable revision (*7*). Interactions between the individual components of massive binary systems are now believed to both be very common and have profound implications for stellar life. In tight binary systems, large amounts of mass can spill over from one object to the other, substantially altering the mass of both stars. Because mass is the overarching parameter controlling evolution, the life cycles and final fates of these interacting stars can deviate considerably from the classical scenario. The net results of such binary interactions are diverse (*8*) and depend on quantities such as the initial separation and masses of the interacting stars as well as their internal, surface, and outflow structures.

One of the most robust predictions of the influence of a close binary companion on massive star evolution is the existence of a population of stars that have been stripped of their hydrogen-rich envelopes by the nearby companion (*8*). Although stripped stars have been observed at the uppermost end of the mass spectrum (*9*), the expected large population of objects produced from stars with 8 to 25 solar masses at birth has never been observed. This is not a trivial issue because it is in this mass range that one expects to find most progenitors of hydrogen-deficient supernovae, binaries emitting fiercely in high-energy x-rays, and gravitational wave mergers involving

Institute of Astrophysics, KU Leuven, Celestijnlaan 200D, 3001 Leuven, Belgium. Email: jon.sundqvist@kuleuven.be; hugues.sana@kuleuven.be

neutron stars. The question has thus been whether this lack of detection points toward some fundamental flaw in the theoretical predictions or whether the stars are simply difficult to detect.

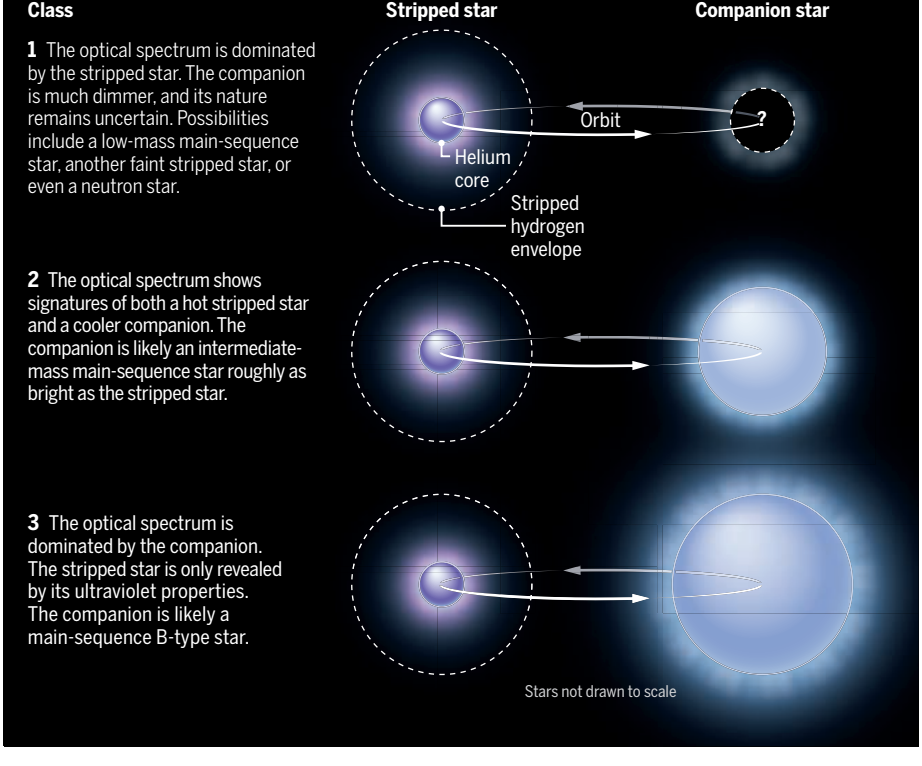
Before the study of Drout *et al.*, only one candidate existed in this stellar mass range. However, the peculiar spectral features of that star were shown to be linked to the presence of an exceptionally strong magnetic field (*10*). A handful of inflated, recently stripped stars have also been identified, but these have not yet reached their hot and compact, long-lived phase (*11*). Thus, these previous detections could not be used as general verification of the missing population. Using ultraviolet (UV) observations of the Magellanic Clouds, two small satellite galaxies to our Milky Way, Drout *et al.* developed new diagnostics to separate the hot and faint stripped stars from the rest of the stellar population. They identified 26 candidates from a sample of more than 500,000. The authors then obtained optical data for the candidates, sorting them into three different categories (see the figure).

Götberg *et al.* used spectral line diagnostics to confirm the predicted hot surface temperature, high surface gravity, and low hydrogen content. Moreover, the authors showed that the optical spectra of stripped stars in this range are dominated by lines in absorption, in stark contrast to the emission line spectra of known stripped stars with higher masses and luminosities (*9*). The reason for this shift in spectral appearance is that for objects with very high luminosity, the force exerted on the gas by stellar photons is so high that a strong wind outflow is launched beneath the visible surface; all we observe from the outside are emission lines originating in the fast, outflowing gas. By contrast, the stripped stars detected by Drout *et al.* have much lower luminosities, resulting in weaker winds launched from above the surface (*12*). Consequently, the optical line spectra of these intermediate-mass stripped stars shift to being mostly in absorption, and this is likely one of the reasons why they have been so difficult to identify. Moreover, because the winds are weak, the masses of these stripped stars are not expected to decrease much more, and their orbits not expected to expand. This suggests that most of them will collapse into neutron stars or black holes, making them excellent progenitor candidates for hydrogen-poor supernovae and gravitational wave mergers (*13*).

The discovery of stripped stars in the Magellanic Clouds paves the way for more observational work. For example, a key next step will be to constrain the orbital properties of the detected stars. This information, combined with more detailed characterization of the physical parameters (mass, temperature, luminosity, mass loss) of the stripped stars, should indicate the type of binary interaction that led to the envelope stripping of the three different categories of binary systems. Furthermore, thousands of stripped stars are predicted to exist in the Magellanic Clouds and the Milky Way (*14*), making it reasonable to expect that the objects identified by Drout *et al.* are only the tip of the iceberg. Challenges in detecting the expected larger populations include obtaining sharper images from denser regions in the sky (where more stars may reside) and gaining a better understanding of interstellar extinction (which can rapidly dim UV stellar light). Additionally, these stars emit most of their radiation in the far UV, but observing facilities that are focused on this electromagnetic window are scarce (and will become scarcer still when the long-standing Hubble Space Telescope is finally decommissioned). Ensuring access to the high-quality observing capabilities required to characterize stripped stars is an important science driver behind the UltraViolet Explorer (UVEX) (*14*), a concept mission under evaluation at NASA. Only with such a facility can stellar binary interaction products be examined systematically, to obtain quantitative constraints on evolution and end-of-life models. This is critical not only for quantifying stellar life now but also for understanding the stellar populations in the early Universe that are now being detected by the JWST. ■

Stripped stars

Three categories of stripped-star binary systems can be distinguished by optical appearance. The optical light is dominated by the stripped star in class 1 systems and by the companion in class 3 systems; in class 2 systems, both stars contribute approximately equally. The nature of the companion star for class 1 systems remains unknown.


Class
Stripped star
Companion star
1 The optical spectrum is dominated by the stripped star. The companion is much dimmer, and its nature remains uncertain. Possibilities include a low-mass main-sequence star, another faint stripped star, or even a neutron star.
Orbit
?
Helium core
Stripped hydrogen envelope
2 The optical spectrum shows signatures of both a hot stripped star and a cooler companion. The companion is likely an intermediate-mass main-sequence star roughly as bright as the stripped star.
3 The optical spectrum is dominated by the companion. The stripped star is only revealed by its ultraviolet properties. The companion is likely a main-sequence B-type star.
Stars not drawn to scale

GRAPHIC: N. BURGESS/*SCIENCE*

REFERENCES AND NOTES

1. H. Sana et al., Astrophys. J. 215 (suppl.), 15 (2014).
2. N. Langer, Annu. Rev. Astron. Astrophys. 50, 107 (2012).
3. M. R. Drout et al., Science 382, 1287 (2023).
4. Y. Götberg et al., arXiv:2307.00074 [astro-ph.SR] (2023).
5. A. Maeder, Physics, Formation, and Evolution of Rotating Stars (Springer, 2009).
6. B. Welch et al., Astrophys. J. 940, 11 (2022).
7. H. Sana et al., Science 337, 444 (2012).
8. P. Marchant, J. Bodensteiner, arXiv:231101865 [astro-ph.SR] (2023).
9. P. Crowther, Annu. Rev. Astron. Astrophys. 45, 177 (2007).
10. T. Shenar et al., Science 381, 761 (2023).
11. J. Villaseñor et al., Mon. Not. R. Astron. Soc. 525, 5121 (2023).
12. N. Moens et al., Astron. Astrophys. 665, A42 (2022).
13. E. Laplace, Y. Götberg, S. E. de Mink, S. Justham, R. Farmer, Astron. Astrophys. 637, A6 (2020).
14. S. Kulkarni et al., arXiv:2111.15608 [astro-ph.GA]

10.1126/science.adl5676

NEURODEGENERATION

Immunotherapies for Alzheimer's disease

Antibodies targeting amyloid-β aggregates slow decline in Alzheimer's disease

By **Todd E. Golde** *and* **Allan I. Levey**

Three monoclonal antibodies—aducanumab, lecanemab, and donanemab—that target amyloid-β (Aβ) deposits in the brain slow cognitive and functional decline in early Alzheimer's disease (AD) (*1*–*3*). The data from the phase 3 clinical trials in mildly symptomatic individuals with AD demonstrate that Aβ reduction has modest clinical benefit. This supports the central tenet of the amyloid cascade hypothesis (ACH), which proposes that Aβ deposits play a causal, initiating role in AD. However, they also show that targeting Aβ deposits in the symptomatic stage of the disease is not a panacea—a finding that is nevertheless consistent with the presumed initiating role of Aβ aggregation in AD pathogenesis. Further, the studies highlight the pivotal role for imaging and fluid biomarkers in developing disease-modifying therapies (DMTs) for AD. Unanswered questions, including those around differences between the treatments, long-term benefits, and the risk-benefit ratio in real-world populations, remain.

For decades, targeting Aβ has been a foundational approach for developing AD DMTs (*4*). After numerous clinical failures with various treatment approaches, supporters of Aβ as a target and of the ACH worried that it might not prove feasible to target Aβ—even if it did play a seminal triggering role in AD. Target engagement and safety issues plagued many previous efforts. Further, many worried that targeting Aβ in the symptomatic phase of AD, in which substantial neurodegeneration is present, might not provide benefit (*5*). Indeed, Aβ aggregation and deposition begins 2 to 3 decades before symptom onset (*6*) and is followed by other pathological changes, including the accumulation of hyperphosphorylated tau protein.

In patients with very early cognitive impairment caused by AD, aducanumab, lecanemab, and donanemab are far from curative, but they all slow cognitive decline by ~25 to 30% over 18 months. The treatments are also associated with marked reductions in biomarkers of AD pathology, including brain amyloid positron emission tomography (PET) and concentrations of phosphorylated tau in cerebrospinal fluid or blood plasma (*2*, *3*). These monoclonal antibodies have some differentiating properties (*7*). However, they share the ability to preferentially bind deposited Aβ and activate immune clearance mechanisms (see the figure). These common properties likely underlie their effects on AD biomarkers and clinical progression as well as their adverse effects. Given the lack of efficacy reported for antibodies with specificity for soluble Aβ, selective targeting of deposited, aggregated Aβ seems to be requisite for efficacy, although targeting of oligomeric assembles could still play a role in treatment. Aducanumab and lecanemab bind Aβ aggregates with a higher degree of affinity compared with monomer. Donanemab targets a pyroglutamine modification in Aβ species that are enriched in Aβ deposits. The trial results for each antibody are more similar than they are different (<10% difference in the rate of slowing), so it is difficult to conclude whether the treatments are truly distinct in terms of clinical effect or whether factors besides the antibody binding properties influenced the relative strength of the clinical trial data. Future studies of these antibodies could reveal whether different biological properties mediate differences in long-term efficacy.

These trials provide strong evidence that reduction of brain Aβ in early symptomatic AD has a positive, but modest, clinical effect. This is bolstered by data from the donanemab trial showing that 47% of treated participants in the earliest disease stage had no clinical progression after 1 year versus 29% on placebo (*2*). However, the notion that other pathologies triggered by Aβ aggregate accumulation become self-reinforcing as symptoms emerge is supported by the observation that on average there is ongoing clinical decline during treatment despite nearly complete elimination of Aβ deposits, as assessed by PET imaging (*8*).

The clinical successes of aducanumab, lecanemab, and donanemab in AD now raise numerous questions. For example, will earlier treatment have a bigger effect? If Aβ deposition is the causal trigger, then clearing Aβ deposits before other pathological features of AD emerge should be an effective prophylactic (preventive) therapy. Data from the donanemab phase 3 study may offer some clues because imaging biomarkers were used to evaluate two groups of participants with confirmed brain Aβ (Aβ positivity)—one with low-to-moderate amounts of pathological tau deposits and another with high amounts of tau deposits (*2*). Treatment showed greater benefit in the group with low-to-moderate tau compared with those with higher tau. These data suggest that treatment in those with more pathological burden is less effective. So, might even earlier treatment be better? Initial prevention trials of lecanemab and donanemab are underway primarily in Aβ-positive, congitively normal individuals.

Most data from the phase 3 trials of aducanumab, lecanemab, and donanemab inform on changes during 18 months of treatment. Several possibilities beyond 18 months can be envisioned. In the best case, the slopes of decline continue to diverge between treatment and placebo groups. Thus an ~25% slowing of decline at 18 months might translate to an ~50% slowing at 3 years. A less promising scenario is that the benefit observed at 18 months is stable, but afterward, decline progresses as in the placebo group. A third scenario is that the effect is transient, and after 18 months, the differentiation between treatment and placebo groups diminishes. Real-world clinical experiences will help to distinguish among these possibilities. In the US, data on individuals receiving the antibodies are being collected in registries.

The overall effects of treatment are important to track, but assessing changes on a more individual basis is important as well. In cancer, many drugs are approved with modest effects on overall survival, but over time, data emerge showing exceptional responses in select individuals who are cured or are in very-long-term remission. So, are there similar responders to these immunotherapies who show very little cognitive decline over extended periods of time? Given the wide variance in the rate of decline in early symptomatic AD, an answer to this question will likely take many years of follow up. Long-term effects and analyses of responses across individuals and subgroups will heavily influence the overall view of the efficacy of these drugs and patient selection.

How safe are these antibodies? All are associated with an increase in abnormal brain magnetic resonance imaging findings. These are referred to as two types of amyloid-related imaging abnormalities (ARIAs): ARIA-E (edema in the brain parenchyma or sulcal effusion—i.e., extravasated fluid in the

Department of Pharmacology and Chemical Biology, Department of Neurology, Center for Neurodegenerative Disease, Goizueta Institute Emory Brain Health, Emory University School of Medicine, Atlanta, GA, USA. Email: tgolde@emory.edu; alevey@emory.edu

leptomeninges along the sulcal spaces) and ARIA-H (hemosiderin deposits resulting from red blood cell breakdown products—i.e., microhemorrhages) (*9*). Although the incidence of ARIAs is increased by antibody treatment and is observed in roughly one-third of treated participants, in most individuals, it is asymptomatic. If symptoms do arise (e.g., in 6% of treated participants in the donanemab phase 3 study), these are typically mild (e.g., headache, confusion, or dizziness). However, in rare cases, more-severe consequences arise (e.g., large hemorrhage in <1% of treated participants), and several deaths have been attributed to antibody-induced severe ARIA (*10*). In all clinical trials of the three antibodies, ARIA was more frequent in the first 6 months of treatment but can occur later, and incidence is approximately three- to sixfold higher in individuals carrying the ϵ4 allele of the apolipoprotein E (*APOE*) gene (*1*–*3*) (the most common genetic risk factor for AD).

What are the biological mechanisms underlying ARIA, and can it be predicted and prevented? Data from trials of antibodies targeting Aβ suggest that ARIA may be a result of antibody engagement of Aβ deposits in the cerebrovascular system and resultant activation of immunologic cascades or perhaps weakening of vessels as Aβ is cleared. Aβ-targeting antibodies that did not bind Aβ deposits or had reduced effector functions did not increase rates of ARIA (*4*). Future studies are needed to understand the biological mechanisms of ARIA and whether it is linked to Aβ clearance.

Is it possible to optimize the clinical implementation of DMTs for AD by using biomarkers and genetic data to identify individuals most likely to benefit considerably from treatment? The answer is yes. Stratification of patients on the basis of disease stage as established through biomarkers and factoring in genetic risk (*APOE*ϵ4 status) provides clear evidence that individualized disease staging affects the overall effect size of the treatments. Such biomarkers also inform on risk for side effects, target engagement, and effects on the brain that may provide hints of long-term benefits and possible unexpected negative properties (e.g., increased rates of brain volume loss). Indeed, phase 2 studies of donanemab clearly showed that inclusion criteria based on imaging biomarker findings (high amyloid and low-to-moderate tau) resulted in a positive trial outcome (*11*). These biomarkers also help to guide dosing and treatment duration, especially in early-phase clinical trials.

Will similar benefits and risks of these anti-Aβ immunotherapies be realized with real-world application? A glaring weakness of many of these trials was the lack of racial diversity among trial participants. Concerted efforts to study these agents in diverse populations using rigorous personalized medicine approaches (i.e., biomarkers) are needed (*12*). Careful monitoring of patients in these trials as well as exclusion criteria designed to limit exposure of individuals potentially at risk of side effects likely helped to limit severe adverse events. As these therapies are given to real-world populations, careful monitoring is needed to ensure that the probability of serious adverse events remains low. Given the modest benefits of treatment, a substantial increase in such events will raise questions about real-world benefit versus risk.

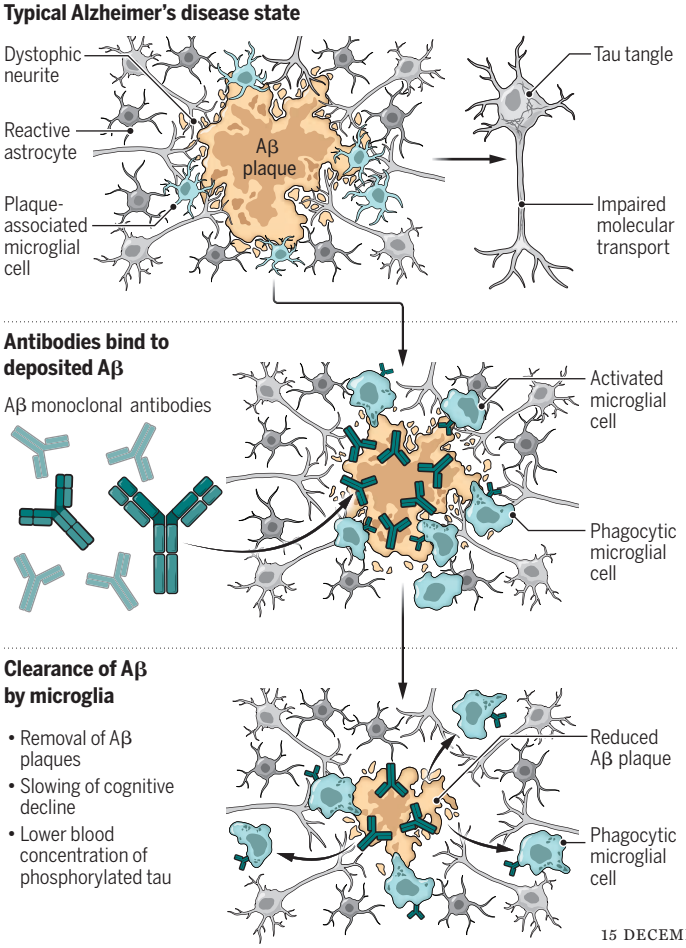
Can treatment be improved? Although these immunotherapies are not panaceas, they are a first step toward the goal of either preventing AD or halting cognitive decline. Better DMTs for AD must be developed. With respect to these or next-generation Aβ immunotherapies, it will be key to develop treatments that are less expensive, safer, more scalable, and more accessible. In the US, aducanumab, lecanemab, and donanemab each have an annual cost of $26,500, but associated imaging and monitoring means that costs for a year of treatment may exceed $75,000. All three monoclonal antibodies are administered through intravenous infusions. Therefore, a pragmatic advance would be to improve the potency so that immunotherapies can be delivered subcutaneously or require fewer infusions, which would reduce cost and thereby increase access.

Longitudinal, omic-based exploratory biomarker studies along with rigorous postmortem analyses of the brains of individuals after treatment could provide a clearer picture of the biology underlying the benefit and risk of treatment. Although supported by some evidence, the concept that deposited Aβ target engagement by an immune-activating antibody results in microglial clearance has not been definitively proven. Obtaining precise mechanistic data from humans could guide the development of an immunotherapy that does not induce ARIA but still shows clinical benefit. Indeed, it is still not known at a detailed pathological and omic level what happens to the brain when amyloid is reduced by these immunotherapies (*13*) or even the true extent to which deposited Aβ is reduced when assessed pathologically and biochemically in post-mortem brain tissue.

These positive data and the need for global scaling should engender rethinking about active vaccines targeting Aβ deposits (*4*). The first trial of an active vaccine for AD (AN-1792) was halted because it resulted in severe meningoencephalitis in

Antibodies mediate removal of amyloid-β

In Alzheimer's disease, deposition of amyloid-β (Aβ) triggers a complex cascade of events that includes the development of tau pathology and results in neurodegeneration and cognitive impairment. Targeting Aβ with monoclonal antibodies decreases brain Aβ concentrations, as assessed by positron emission tomography, and modestly slows cognitive decline in individuals with early symptomatic Alzheimer's disease. The antibodies also lower amounts of phosphorylated tau in the blood (a biomarker of Alzheimer's disease). At a mechanistic level, the antibodies are thought to induce clearance of plaques by microglial cells, although this has yet to be confirmed.



GRAPHIC: A. FISHER/*SCIENCE*

~6% of participants, and only 20% of participants showed a strong antibody response to Aβ. Might it be possible to generate a vaccine that can induce a robust antibody response that specifically targets deposited Aβ without causing severe side effects (*7*)? Efforts are ongoing to develop active vaccines targeting Aβ. Although several of these vaccines currently in trials appear to be safe, it is unclear whether they robustly induce the production of antibodies that can selectively target deposited Aβ. In any case, a safe and effective vaccine could be used prophylactically at a fraction of the cost of monoclonal antibodies.

More generally, the data speak to the need for non–amyloid-targeting DMTs and better therapies for the symptoms of the disease. More substantial clinical benefits might be anticipated by treatments that ameliorate other pathophysiological processes involved in AD, such as neurofibrillary tangle formation, synapse loss, neuroinflammation, metabolic failure, and cell death. Even if one day AD is preventable, millions of people will develop symptomatic AD before then (*14*). Therefore, interventions, perhaps even combination therapies, that more substantially alter disease progression must be developed.

Many factors contribute to cognitive resilience and cognitive decline in late life, some of which may have little to do with the hallmark pathologies of AD (*8*, *15*). Given this complexity, biomarker-guided personalized medicine approaches will likely be paramount for the development of the next generations of AD therapies and, more broadly, dementia therapies. Hopefully the lessons learned from previous failures and successes for AD DMTs can be pragmatically applied to streamline therapeutic discovery, development, and clinical testing. ■

REFERENCES AND NOTES

1. S. Budd Haeberlein *et al.*, *J. Prev. Alzheimers Dis.* **9**, 197 (2022).
2. J. R. Sims *et al.*, *JAMA* **330**, 512 (2023).
3. C. H. van Dyck *et al.*, *N. Engl. J. Med.* **388**, 9 (2023).
4. T. E. Golde, *Neurotherapeutics* **19**, 209 (2022).
5. E. McDade *et al.*, *Mol. Neurodegeneration* **16**, 49 (2021).
6. H. Hampel *et al.*, *Nat. Rev. Neurol.* **17**, 580 (2021).
7. L. Söderberg *et al.*, *Neurotherapeutics* **20**, 195 (2023).
8. T. E. Golde, *Mol. Neurodegeneration* **17**, 18 (2022).
9. R. A. Sperling *et al.*, *Alzheimers Dement.* **7**, 367 (2011).
10. N. J. Reish *et al.*, *N. Engl. J. Med.* **388**, 478 (2023).
11. M. A. Mintun *et al.*, *N. Engl. J. Med.* **384**, 1691 (2021).
12. J. J. Manly, K. D. Deters, *JAMA* **330**, 510 (2023).
13. E. C. B. Johnson *et al.*, *Nat. Med.* **26**, 769 (2020).
14. T. E. Golde *et al.*, *Science* **362**, 1250 (2018).
15. M. A. DeTure, D. W. Dickson, *Mol. Neurodegener.* **14**, 32 (2019).

ACKNOWLEDGMENTS

T.E.G. is an inventor on multiple patents related to Alzheimer's disease therapeutics and is a cofounder of Lacerta Therapeutics and Andante Biologics. A.I.L. has received research support for clinical studies of aducanumab and lecanemab; is a cofounder of and consultant for EmTheraPro; and is a consultant for Cognito Therapeutics, Karuna Therapeutics, Linus Health, MEPSGEN, and NextSense.

10.1126/science.adj9255

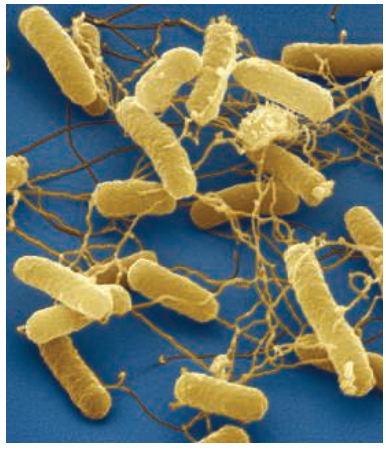
MICROBIOLOGY

No room at the table

The gut microbiota prevents infection by crowding out pathogens

By **Lauren C. Radlinski** *and* **Andreas J. Bäumler**

A vital role of the gut microbiota is to defend the host from infection with pathogens through a process called colonization resistance. For example, mice raised without a microbiota are viable in their sterile isolators but rapidly succumb to infection when exposed to the outside world. Similarly, disruption of gut microbiota in patients through treatment with antibiotics incurs the risk of subsequent life-threatening opportunistic infection. Despite its importance, what constitutes a protective microbiota and the types of communities involved are poorly understood. On page 1259 of this issue, Spragge *et al.* (*1*) take a systematic approach to determine the functional characteristics of a protective microbiota, and apply these findings towardpredicting bacterial communities that guard against important human pathogens. As a result, the authors integrate decades of descriptive studies into a mechanistic understanding of how microbe-microbe interactions influence human health. This could guide the rational design of probiotic communities that confer protection against opportunistic infection.



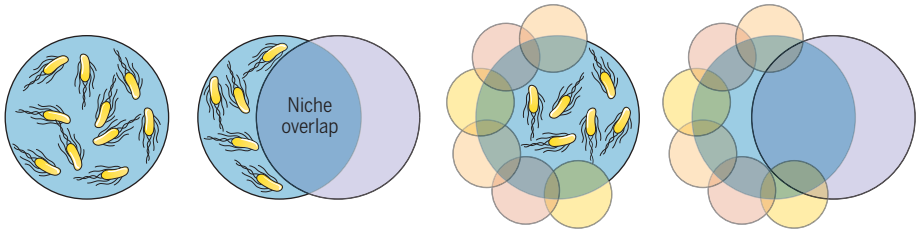
Colonization of pathogens, such as *Salmonella enterica* serovar Typhimurium, may be resisted by microbiota that consume the same nutrients.

There has been a substantial effort to associate microbial community composition with health and disease. Yet, despite the abundance of data generated by advancements in sequencing technologies, it remains unclear what constitutes a "healthy" microbiota or how to define an impaired one. This information would make it possible to predict susceptibility to opportunistic infection by metagenomic analysis of patient feces. Studies have long hinted at a correlation between microbial community diversity and host health, without a clear explanation as to why (*2*). Spragge *et al.* sought to identify the underlying ecological principles behind these observations, rather than focusing on specific bacterial species, which may differ amongindividuals. Through in silico and in vitro analysis of carbon sources used by individual members of defined microbial consortia, they determined that complex communities confer colonization resistance by depleting all the resources required for growth of an opportunistic pathogen, such as *Klebsiella pneumoniae*, or a pathogen, such as *Salmonella enterica* serovar Typhimurium. The overlap of nutritional requirements was greatest between these pathogens and related species within the microbial consortia, such as *Escherichia coli*, thus making the latter a keystone species for colonization resistance. However, this keystone species was only important in the context of a diverse community. This suggests that colonization resistance results from complex, "higher-order" effects that occur when interspecies interactions are influenced by the larger microbial community (see the figure).

Current probiotics are often simplified combinations of species that have a loose association with health. The efficacy of these probiotic interventions after antibiotic therapy are controversial, because they can interfere with microbiota reconstitution (*3*, *4*). Spragge *et al.* show that the nutritional requirements for pathogens and probiotic species are often very different, and in cases where competitors share similar nutritional niches, probiotic-pathogen competition is only protective in the context of a diverse community. The authors mined existing repositories of human-derived microbial genomic sequences to identify communities with sequence overlap to relevant pathogens (such as *E. coli*) and accurately predict species that curb pathogen growth. This suggests that probiotic communities

Resisting colonization by pathogens

Colonization resistance against opportunistic pathogens is mediated by a diverse microbiota community that has substantial niche overlap. Without the protection of a microbiota, opportunistic pathogens grow unhindered within the host (**1**). Individually, the presence of a single keystone species that has considerable nutritional overlap with the pathogen (**2**) or a diverse community of unrelated commensal species (**3**) is insufficient to curb pathogen growth. However, interspecies competition between related species in the presence of a diverse community is sufficient to confer colonization resistance to protect the host from infection with opportunistic pathogens (**4**).


Niche overlap
1 Pathogen
2 Pathogen + keystone species
3 Pathogen + diverse community
4 Keystone species + diverse community

GRAPHIC: K. HOLOSKI/*SCIENCE*

can be tailored to protect patients at high risk of certain infections, for instance, after completing antibiotic therapy in a hospital.

An interesting observation from the study of Spragge *et al.* was that supplementing mouse feed with resources that are only accessible to the pathogen, such as the sugar alcohol galactitol, could circumvent an otherwise protective microbiota by providing a nutrient only the pathogen could use. This aligns with recent reports that sugar alcohols fuel the bloom of pathogenic organisms, even in the presence of an intact microbiota (*5*) and is of particular interest considering the recent rise in popularity of sugar alcohols as noncaloric substitutes for simple sugars. Sugar alcohols, such as mannitol, isomalt, and sorbitol, are sweeteners that reduce the caloric content of food, but may come at the cost of providing pathogens a fitness advantage to invade and disrupt an otherwise stable community.

Similarly, pathogens such as *Salmonella* actively modulate the host environment during the course of infection because intestinal inflammation generates new carbon sources in the gut (*6-8*). Furthermore, when *Salmonella* induces gastrointestinal inflammation, the availability of oxygen and nitrate increases and permits the pathogen to gain a competitive advantage over resident bacteria through aerobic respiration (*9*). A limitation of the study by Spragge *et al.* is that by using an avirulent mutant of *Salmonella* to infect mice, the authors do not account for an increased availability of respiratory electron acceptors, such as oxygen and nitrate, which changes the redox reactions that can be used to catabolize carbon sources during *Salmonella*-induced gut inflammation. These changes would likely not be accounted for during the Biolog plate assays the authors performed, but may show up in the sequence-based analysis for identifying metabolic overlap. *E. coli,* for instance, shares several oxygen- and nitrate-dependent metabolic pathways with *Salmonella,* which have been shown to be important for colonization resistance against *Salmonella* during host inflammation (*10-12*). Further investigation is required to determine how conclusions about opportunistic pathogens, such as *K. pneumoniae*, can be extrapolated to pathogens, such as *Salmonella*, which use their virulence factors for ecosystem engineering.

Spragge *et al.* demonstrate that it is possible to apply simple ecological principles to gain a better understanding of the interspecies interactions that take place within the gut microbiome. Moving forward, it will be interesting to see if these governing principles hold for other tissues of the body that house a microbiome, as well as how host dynamics, such as inflammatory responses, influence these higher-order interactions. Given widespread antibiotic resistance, improved predictive tools such as that developed by Spragge *et al.* could allow clinicians to administer tailored probiotic communities that prevent difficult or impossible-to-treat infections from occurring in the first place. ■

Department of Medical Microbiology and Immunology, School of Medicine, University of California Davis, Davis, CA, USA. Email: ajbaumler@ucdavis.edu

REFERENCES AND NOTES

1. F. Spragge *et al.*, *Science* **382**, eadj3502 (2023).
2. O. Manor *et al.*, *Nat. Commun.* **11**, 5206 (2020).
3. G. T. Rijkers *et al.*, *Br. J. Nutr.* **106**, 1291 (2011).
4. J. Suez *et al.*, *Cell* **174**, 1406 (2018).
5. C. Eberl *et al.*, *Cell Host Microbe* **29**, 680 (2021).
6. S. E. Winter *et al.*, *Nature* **467**, 426 (2010).
7. F. Faber *et al.*, *Nature* **534**, 697 (2016).
8. C. C. Gillis *et al.*, *Cell Host Microbe* **23**, 54 (2018).
9. F. Rivera-Chavez *et al.*, *Microbe* **19**, 443 (2016).
10. E. M. Velazquez *et al.*, *Nat. Microbiol.* **4**, 1057 (2019).
11. Y. Litvak *et al.*, *Cell Host Microbe* **25**, 128 (2019).
12. M. J. Liou *et al.*, *Cell Host Microbe* **30**, 836 (2022).

10.1126/science.adl5891

EVOLUTION

Tracing the origin of virulence

Microbial genomes from ancient chickens uncover the drivers of pathogenicity

By **Sebastian Duchene**[1,2]

Analyzing the genomes of ancient pathogens has revolutionized the understanding of how they originate and spread. Most of this work has concentrated on human pathogens, but there is an increasing need to explain the origin of infectious disease in wildlife and domesticated animals (*1*). On page 1276 of this issue, Fiddaman *et al.* (*2*) obtain the genomes of Marek's disease virus (MDV) from chicken remains from as long ago as the 10th century. This virus was described in the early 1900s as the causative agent of a mild tumor disease in poultry, but from the 1950s, the virus was found to produce very high mortality rates in unvaccinated chickens. The authors found that MDV has circulated in poultry for at least 1000 years but only acquired genetic changes for increased virulence in the past century. The work provides vital information on how pathogens persist in host populations and how ecology can drive pathogenicity.

The past decade has seen a surge in the genomic sequencing of ancient microbes, largely owing to advances in sequencing technologies (*3*, *4*). Phylogenetic analyses can position ancient microbes in an evolutionary and chronological context with respect to modern pathogens (*5*, *6*). The recovery of genomes from *Mycobacterium tuberculosis,* the bacterium responsible for tuberculosis, from precolonial mummies revealed that the modern strains and those that circulated in humans in precolonial America are intermingled with strains that infect a range of mammals, including some pinnipeds and ungulates. This observation indicates a potential role for repeated zoonotic spillovers (in which animal pathogens become capable of infecting humans) (*7*).

[1]Department of Computational Biology, Institut Pasteur, Paris, France. [2]Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia. Email: sduchene@pasteur.fr

An increase in virulence of Marek's disease coincided with the growth of domesticated chicken populations.

Obtaining ancient genomes from viruses is generally more challenging than for bacteria and multicellular organisms because viral genetic material can more easily degrade and is of lower abundance (*3*). However, numerous recent studies have generated such data for viruses with both DNA and RNA genomes (*4*). In 2018, analysis of hepatitis B viral genomes from humans from 400 to 10,500 years ago supported a long-term association between this pathogen and humans, with some lineages of the virus present in human populations for at least 11,000 years (*8*). Similarly, the sequencing of viral samples from the 1918 to 1920 influenza pandemic revealed an avian origin of the virus with key adaptations to mammalian hosts (*6*).

Determining the timescale of evolution of many pathogens has only been possible through the analysis of ancient samples. However, an emerging question that warrants attention is when and how virulence, the ability of an organism to infect a host and cause disease, is shaped over time. Ancient genomics allows genomic regions from modern strains that are responsible for infection or pathogenesis to be compared with those of ancient strains. The main requirement is that the quality of the ancient genome must be high enough to enable the identification of such genes or molecular markers. The first step in identifying the evolution of virulence is to compare genomic content. For example, modern pathogen genomes may have a set of genes associated with virulence that is absent in their ancient counterparts. In many bacteria, pathogenesis is driven by the presence or absence of large portions of genetic material (*9*). Fiddaman *et al.* found that for MDV, the gene content of modern and ancient strains is very similar, and virulence appears to be driven by specific mutations in genes that are associated with oncogenesis.

A notable finding in the study of several modern and ancient outbreaks of infectious disease is that mutations that confer increased virulence or transmissibility accumulate faster than other mutations. For example, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern display mutations that affect vaccine efficacy and increase virus transmissibility (*10*). The lineages that gave rise to variants of concern were shown to accumulate mutations fourfold faster than other lineages of the virus (*11*). Similarly, the Mpox virus outbreak detected in 2022, appears to have rapidly accumulated mutations that allow it to transmit more effectively between humans (*12*). Bacterial outbreaks of plague (*Yersinia pestis*) and leprosy (*Mycobacterium leprae*) also display evidence of rapid accumulation of virulence mutations, a finding that has been supported by analysis of ancient molecular data (*13*, *14*).

An ecological perspective is key to understanding the conditions that favor the accumulation of mutations that confer increased virulence. Molecular clock models, part of the phylogenetic toolkit, are commonly used to estimate the timing of evolutionary events. The accuracy of these models vastly improves with the inclusion of ancient genomes. Fidamann *et al.* used a molecular clock to determine that the mutations that characterize the high virulence of MDV present today appeared around the time when poultry farming was rapidly expanding, reflecting a period of tremendous ecological transition. The recent practice of sustaining large and dense poultry populations can facilitate the spread of a highly virulent pathogen. By contrast, in ancient populations that tended to be smaller and more isolated, such a lethal pathogen would have been more likely to rapidly kill its hosts and become locally extinct, preventing it from spreading more widely and therefore posing a fitness cost to high virulence. Moreover, high-density living conditions of the host can reduce the time between transmission events, increasing the rate at which mutations can appear in the pathogen.

The emergence and persistence of pathogens are determined by the interplay of ecological and immunological barriers (*15*). Several major trends in ancient genomics will play important roles in improving knowledge of this relationship. Notably, the study of pathogens of domesticated and wild species will help establish the role of ecology in disease emergence.

An increase in the number and quality of ancient genomes is a promising avenue for understanding the origin and evolution of pathogenicity. As a case in point, the appearance of many zoonotic virus outbreaks in humans—for example, SARS-CoV-2, Ebolavirus, Mpox virus, and Middle East respiratory syndrome coronavirus (MERS-CoV)—regularly prompts questions as to whether their virulence will increase or decrease over time as they spread in a new host species. The emergence of SARS-CoV-2 variants showed that this is difficult to determine and that interactions between host immunity and behavior can shape the evolution of viruses. Ancient genomes of pathogens with long-term associations with humans and other animals, such as herpes viruses and hepatitis B virus, will be enormously useful to elucidate the interaction between emergence, pathogenicity, and ultimately the timescale over which they become established in their hosts. ■

REFERENCES AND NOTES

1. E. C. Holmes, *Annu. Rev. Virol.* **9**, 173 (2022).
2. S. R. Fiddaman *et al.*, *Science* **382**, 1276 (2023).
3. M. A. Spyrou, K. I. Bos, A. Herbig, J. Krause, *Nat. Rev. Genet.* **20**, 323 (2019).
4. S. Duchêne, S. Y. W. Ho, A. G. Carmichael, E. C. Holmes, H. Poinar, *Curr. Biol.* **30**, R1215 (2020).
5. S. Y. W. Ho, S. Duchêne, *Science* **368**, 1310 (2020).
6. A. A. Guzmán-Solís, M. A. Navarro, M. C. Ávila-Arcos, D. Blanco-Melo, *Annu. Rev. Virol.* **10**, 49 (2023).
7. Å. J. Vågene *et al.*, *Nat. Commun.* **13**, 1195 (2022).
8. A. Kocher *et al.*, *Science* **374**, 182 (2021).
9. R. J. Bennett, K. S. Baker, *J. Clin. Microbiol.* **57**, 00100-19 (2019).
10. D. P. Martin *et al.* NGS-SA; COVID-19 Genomics UK (COG-UK), *Cell* **184**, 5189 (2021).
11. J. H. Tay, A. F. Porter, W. Wirth, S. Duchene, *Mol. Biol. Evol.* **39**, msac013 (2022).
12. Á. O'Toole *et al.*, *Science* **382**, 595 (2023).
13. J. Klunk *et al.*, *Nature* **611**, 312 (2022).
14. A. Benjak *et al.*, *Nat. Commun.* **9**, 352 (2018).
15. R. K. Plowright *et al.*, *Nat. Rev. Microbiol.* **15**, 502 (2017).

ACKNOWLEDGMENTS

S.D. was funded by Institut Pasteur and the National Health and Medical Research Council of Australia.

10.1126/science.adl6094

MATERIALS SCIENCE

Solar-powered clothes, for the heat and cold

A wearable device allows the human body to adapt to changes in ambient temperature

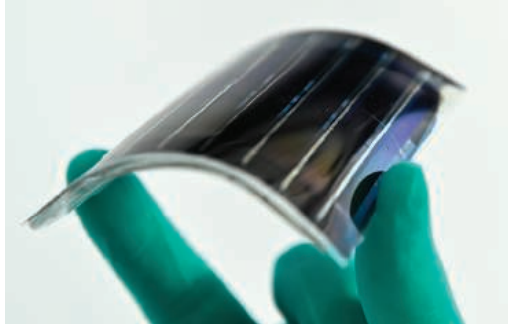
By **Xingyi Huang**[1,2] *and* **Pengli Li**[1]

Clothing plays an indispensable role in maintaining the human body temperature within a comfort range in our daily life, especially when facing sudden temperature changes or in harsh environments (*1*). Failure to maintain core body temperature during complex environmental changes can result in severe medical emergencies (*2*). On page 1291 of this issue, Wang *et al.* (*3*) report a full-day, self-powered, and bidirectional thermoregulatory clothing that can quickly respond to fluctuating temperature. This thermoregulatory clothing could guarantee the safety and comfort of the human body amid fluctuating environmental temperatures and even extend human survivability in extreme environments, such as polar regions and outer space.

Thermoregulatory clothing can be classified into two categories. The passive type encompasses radiative cooling, phase change, and adsorption systems, which have the advantage of requiring no external energy inputs. The active type can facilitate on-demand rapid cooling or warming of the human body. These include solid-state active thermoregulation systems that are based on the Joule heating effect (*4*) and the Peltier effect (*5*). However, because of high energy consumption, it is difficult to maintain continuous thermoregulation of the human body for a long time in such systems through portable, sustainable energy-harvesting devices. In recent years, flexible organic photovoltaic cells that use sunlight have made substantial progress, not only achieving nearly 20% of photovoltaic conversion efficiency (solar light into electrical energy) (*6*) but also enabling large-scale production (*7*, *8*). Thus, if such a highly efficient and flexible organic photovoltaic unit could be integrated with a proper thermal-management system, then robust, self-sustaining, and thermoregulating clothing could be achieved.

Since the discovery of polyvinylidene fluoride–based ferroelectric polymers, many advances have been made in the performance and structure of electrocaloric devices—objects that exhibit a reversible temperature change in response to an applied electric field (*9*–*11*). However, the energy source for actuating an electrocaloric device typically uses grid electricity, and the structure of the device may be large, thus limiting its integration into clothing. Additionally, a wearable electrocaloric device must be flexible (*12*).

Wang *et al.* accounted for these factors and effectively designed and fabricated a wearable thermal-management system by combining an organic photovoltaic unit and an electrocaloric unit into a single device with the required flexibility. The device also achieved bidirectional thermal management, providing 10.1 K of cooling to the skin during hot days but also keeping the human body 3.2 K warmer than bare skin in the dark or at night by using additional energy collected by the organic photovoltaic unit. Moreover, skin could be warmed at a maximum rate of 15.6°C/min or cooled at a maximum rate of 14.0°C/min in the first 5 s to achieve fast thermoregulation. With its excellent thermal-management performance, easy switching of the thermal-management direction, and adjustment of the thermal-management temperature range for optimal temperature control, Wang *et al.* have demonstrated clothing that allows the human body to adapt to changes in ambient temperature.



A self-powered, wearable thermoregulatory system provides efficient personalized thermoregulation, extending the human thermal comfort zone by 19.1 K.

The thermal comfort zone for the human body is mainly based on the range of ambient temperatures. The upper limit is the temperature at which the human body feels "hot," and the lower limit is the temperature at which the human body feels "cold." For bare skin, this ranges from 22° to 23°C to 27° to 28°C (the comfort zone is thus 4 to 6 K). Previous work had achieved an expansion of the thermal comfort zone, from 4 to 10 K, by flipping clothing inside out to tune thermal radiation (*13*). Wang *et al.* expanded the comfort zone by an impressive 19.1 K. Such an increase should presumably allow the human body to adapt to more complex and changing environments. However, this system still needs to be optimized in the future in terms of performance and practicality, including the temperature span of the system and the thermal conductivity of the materials. The device of Wang *et al.* could be further improved by increasing the entropy change of the materials, and the thermal conductivity of a material can be improved by adding inorganic fillers (*14*). In addition, the thermal-management performance of the whole system could be optimized by constructing a cascade device in which multiple electrocaloric devices are sequentially linked and operate in synergy to increase the temperature span of the whole system.

The integrated device of Wang *et al.* opens many possibilities for developing actively controlled, self-powered, and wearable localized thermal-management systems and expanding human adaptation to harsh environments. Beyond clothing, such devices could be applied to vehicles and buildings. It is possible to imagine a future of all-weather thermal management that is not limited by an energy supply and where extra collected energy might even power electronic devices under special conditions. ■

[1]Shanghai Key Laboratory of Electrical Insulation and Thermal Aging, School of Chemistry and Chemical Engineering, Frontiers Science Center for Transformative Molecules, Shanghai Jiao Tong University, Shanghai, China. [2]Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. Email: xyhuang@sjtu.edu.cn

PHOTO: X. HUANG AND P. LI

REFERENCES AND NOTES

1. J. Chai *et al.*, *Cell Rep. Phys. Sci.* **3**, 100958 (2022).
2. ANSI/ASHRAE, *Standard 55-2017: Thermal Environmental Conditions for Human Occupancy* (ASHRAE, 2017).
3. Z. Wang *et al.*, *Science* **382**, 1291 (2023).
4. Y. Guo *et al.*, *Small* **13**, 1702645 (2017).
5. H. Wei *et al.*, *Appl. Energy* **326**, 119941 (2022).
6. J. Wang *et al.*, *Natl. Sci. Rev.* **10**, nwad085 (2023).
7. H. Jinno *et al.*, *Nat. Energy* **2**, 780 (2017).
8. Y.-F. Shen *et al.*, *Adv. Mater.* **35**, 2209030 (2023).
9. X. Qian *et al.*, *Nature* **600**, 664 (2021).
10. A. Torelló *et al.*, *Science* **370**, 125 (2020).
11. Y. Wang *et al.*, *Science* **370**, 129 (2020).
12. R. Ma *et al.*, *Science* **357**, 1130 (2017).
13. P.-C. Hsu *et al.*, *Sci. Adv.* **3**, e1700895 (2017).
14. P. Li *et al.*, *Adv. Funct. Mater.* **32**, 2109542 (2022).

ACKNOWLEDGMENTS

X.H. was supported by National Natural Science Foundation of China grant 51877132.

10.1126/science.adl5650

POLICY FORUM

BIODIVERSITY

Securing success for the Nature Restoration Laws

The EU law would complement many others, but challenges loom

By **Daniel Hering[1,2], Christian Schürings[1], Franziska Wenskus[1], Kirsty Blackstock[3], Angel Borja[4], Sebastian Birk[1,2], Craig Bullock[5,6], Laurence Carvalho[7], Magda Bou Dagher-Kharrat[8], Sebastian Lakner[8], Nataša Lovrić[9,10], Shane McGuinness[5,6], Gert-Jan Nabuurs[11], Agustín Sánchez-Arcilla[12], Josef Settele[13,14,15], Guy Pe'er[14,16]**

In an attempt to halt and reverse biodiversity losses, the European Commission has proposed a new regulation, the Nature Restoration Law (NRL). It could become a cornerstone of Europe's ambitions to restore biodiversity and ecosystem services for decades to come (*1*) and demonstrate global leadership in addressing ongoing environmental crises. The draft of the law, which is a first globally, has been under political pressure from various sides, and scientists have contributed intensively to the discussion (*2*). After trilogue negotiations among the European Parliament, the Council of Europe, and the European Commission, the final text of the NRL has been agreed on (see the box). However, it will still be subject to final votes within the Council and Parliament. Here, we assess the potential for the NRL to overcome problems associated with implementation of related European Union (EU) legislation, strategies, and policies and what can be learned for implementation of the NRL.

The NRL acknowledges that existing EU legislation and policies have so far failed to halt biodiversity losses (*1*) and consequently, without new instruments, cannot meet the targets of international agreements, such as the Kunming-Montreal Global Biodiversity Framework. Although some of the NRL's aims and approaches overlap with other EU directives, strategies, and policies, in particular with the EU Biodiversity Strategy for 2030, the NRL is distinct in terms of its coverage targeting the majority of European ecosystem types, its strong focus on restoration, and its provision of binding targets and clear timelines. This potential for regulatory power may largely explain the contested nature of its passage into legislation.

The prospect of the NRL achieving its aims will be strongly determined by other European legislation and policies that address the environment as well as land and water uses (see fig. S1). Policy coherence requires complementary objectives and instrument mixes within environmental domains (*3*) while mainstreaming environmental objectives into other policy domains (*4*). These may enhance options for, or pose restrictions on, the implementation of the NRL. Key directives, some of which came into force decades ago, include the Habitats Directive (HD), Birds Directive (BD), Water Framework Directive (WFD), and Marine Strategy Framework Directive (MSFD). They share aims in safeguarding Europe's biodiversity but have not halted its decline. The Biodiversity Strategy for 2030 (BS) targets halting biodiversity loss, while the Forest Strategy (FS) and Common Fisheries Policy (CFP) address major land and sea uses. Last, the Common Agricultural Policy (CAP) has the largest budget and affects nearly 40% of the EU's terrestrial area, yet agriculture remains the lead driver of biodiversity loss (*5*). Together, these directives and policies cover a broad range of targets, sectors, and approaches and are representative for other instruments that will also interact with the NRL implementation (see the supplementary materials for details on our analysis of existing legislation, strategies, and policies).

"This potential for regulatory power may largely explain the contested nature of its passage into legislation."

LESSONS LEARNED

In developing the NRL, the EU has learned from past experiences with European environmental legislation and policies and avoids several obstacles that have obstructed their implementation.

As a regulation, the NRL will come into force soon after it has been passed by the EU Parliament. This is an advantage in comparison with the HD/BD, WFD, and MSFD, which needed to be transposed into national law—a process that takes several years. Although the NRL will also need national implementation—for example, through National Restoration Plans (NRPs)—these could be passed by authorities without legislative procedures. This is a major advantage because speed is vital for tackling the biodiversity crisis and fulfilling the EU's international commitments (*6*).

The NRL sets ambitious quantitative targets in terms of both the areas to restore and the timeframe, with targets for 2030, 2040, and 2050 (see the box). Experiences with previous legislation support this approach. The WFD and the MSFD defined deadlines for meeting the good status of all water bodies and seas (although in the case of the WFD allowing for an extension), but these firm deadlines made continuous restoration activities with intermediate targets more difficult. Timing, however, is

[1]Department of Aquatic Ecology, University of Duisburg-Essen, Essen, Germany. [2]Centre of Water and Environmental Research, University of Duisburg-Essen, Essen, Germany. [3]Social, Economic and Geographical Sciences, James Hutton Institute, Aberdeen, UK. [4]AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Pasaia, Spain. [5]School of Architecture, Planning and Environmental Policy, University College Dublin, Dublin, Ireland. [6]Earth Institute, University College Dublin, Dublin, Ireland. [7]Freshwater Ecology Section, Norwegian Institute for Water Research (NIVA), Oslo, Norway. [8]Faculty of Agricultural and Environmental Sciences, Chair of Agricultural Economics, University of Rostock, Rostock, Germany. [9]European Forest Institute, Joensuu, Finland. [10]School of Forest Sciences, University of Eastern Finland, Joensuu, Finland. [11]European Forest Resources, Wageningen University and Research, Wageningen, Netherlands. [12]Maritime Engineering Laboratory, Escola de Camins, Department of Civil and Environmental Engneering, UPC BarcelonaTech, Barcelona, Spain. [13]UFZ–Helmholtz Centre for Environmental Research, Department of Conservation Biology and Social-Ecological Systems, Halle, Germany. [14]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. [15]Institute of Biological Sciences, University of the Philippines Los Banos, College, Laguna, Philippines. [16]UFZ–Helmholtz Centre for Environmental Research, Department of Ecosystem Services, Leipzig, Germany. Email: daniel.hering@uni-due.de

also an issue for the NRL. The NRL's success hinges on prompt action and the provision of effective tools for achieving targets within short timeframes, recognizing the necessary time for nature to recover.

The NRL defines measurable and applicable indicators for restoration success. These include the area of restored habitats, which is easy to document and to control. Other more generic indicators, such as the Grassland Butterfly Index, are well established, thus facilitating implementation. A third group of indicators will require some standardization, such as indicators of forest restoration. Although no specific indicators are defined for marine ecosystems, criteria from the MSFD could be applied. Hence, the NRL can to a large degree capitalize on existing indicators, in sharp contrast to the WFD, MSFD, and HD/BD, all of which ignited extensive indicator development processes that delayed implementation.

Another advantage is the use of NRPs, which has the potential to provide an appropriately tailored national framework for NRL implementation. Although all the above-listed directives are implemented at the national level, there have been particularly good experiences when actions take account of local contexts and needs, as seen with the River Basin Management Plans under the WFD. It is therefore imperative to ensure that the NRPs will be backed by robust implementation tools that adopt an adaptive cycle, whereby the commission can request member states to increase their ambition.

Key features of the EU Nature Restoration Law

Chapter I: General Provisions

- Defines the overall targets (continuous recovery of nature, fulfillment of climate change objectives and international regulations)
- Defines key terms: Favorable reference area (minimum area to ensure the long-term viability of a habitat type), good condition (characteristics that ensure favorable conservation status according to the HD or good environmental status according to the MSFD), sufficient quality and quantity of habitat (conditions required by a species for maintaining itself on a long-term basis)

Chapter II: Restoration Targets and Obligations

- For Natura 2000 sites (Article 4): Good condition (30% by 2030, 60% by 2040, 90% by 2050); and favorable reference area (30% of the area needed to reach the goal for each habitat type by 2030, 60% by 2040, 100% by 2050); improved connectivity
- For habitats of species listed in Annexes II, IV, and V of Habitats Directive and of Birds Directive (Art. 4): Reach sufficient quality and quantity of habitats (no time frame given)
- Marine Ecosystems (Art. 5): Reaching good condition (30% by 2030, 60% by 2040, 90% by 2050) and favorable reference area (at least 30% by 2030, 60% by 2040, and 100% by 2050)
- Urban Ecosystems (Art. 6): No loss in total national area of urban green spaces, achieve thereafter an increasing trend
- Rivers, floodplains (Art. 7): Removal of barriers to longitudinal and lateral connectivity to achieve restoration targets and 25,000 km of free-flowing rivers; maintain and improve natural functions of floodplains
- Pollinator populations (Art. 8): Improve pollinator diversity, reverse decline of pollinator populations by 2030; achieve thereafter an increasing trend of pollinator populations
- Agricultural ecosystems (Art. 9): Increasing trend at national level in two of the three indicators: "Grassland Butterfly Index," "stock of organic carbon in cropland mineral soils," "share of agricultural land with high-diversity landscape features"; targets for "Common Farmland Bird Index": increase by 10% (2030), 20% (2040), and 30% (2050) for member states with depleted farmland bird populations, and by 5% (2030), 10% (2040), and 15% (2050) for member states with less depleted populations; restoration of organic soils in agricultural use constituting drained peatlands: 30% (by 2030), 40% (by 2040), 50% (by 2050)
- Forest ecosystems (Art. 10): Increasing trend at national level of the "Common Forest Bird Index" and in 6 out of 7 additional indicators such as standing deadwood or forest connectivity

Chapter III: National Restoration Plans

- Obliges member states to prepare restoration plans to implement the measures required for targets of Chapter II, and to quantify the area to be restored
- Member states have full flexibility to use or to discard funds from Common Agricultural Policy and Common Fisheries Policy for NRL implementation

Chapter IV: Monitoring

- Obliges member states to monitor indicators for restoration targets; progress reports by the Commission

Chapter VI: Final Provisions

- Application of the NRL will be evaluated by 2033, including possibly legislative proposals for amendments
- "Emergency brake" allows member states to halt NRL implementation in farmland, if agricultural production is at risk

ADVANCING IMPLEMENTATION

The NRL's aims reach well beyond the targets of existing legislation and policies (see the box). In addition, the NRL offers great potential to boost the implementation of other European directives and policies. Whereas the WFD and MSFD focus on individual ecosystem types (surface water and marine ecosystems), the HD/BD take a broader approach, including a wide range of habitats, and the BS is even more comprehensive (*7*) because it addresses species, habitats, ecosystems, ecological processes, and public engagement. The NRL is broad but targets specific ecosystem types with tailor-made approaches (see the box and fig. S1). It may therefore have impacts beyond the targeted ecosystems: For example, restoring agricultural ecosystems and forests has the potential to benefit rivers and lakes, and restoring peatlands can positively affect the landscape's water budget (*8*). Consequently, implementation of the NRL can substantially benefit the implementation of the HD/BD, WFD, and MSFD. This is most obvious for the HD/BD, which addresses a greatly overlapping list of habitats and species. The WFD and the MSFD can benefit from reduced pollution from agriculture and from the additional approaches the NRL provides. For example, the WFD does not explicitly address floodplains, although floodplains play an important role in the healthy functioning of rivers and their ecological quality (*9*). Also, the implementation of the BS will benefit from the restoration measures initiated by the NRL.

At first glance, the NRL may seem to be "conservative." It focuses mainly on the protection and restoration of habitats per se and of habitats for individual species. This is reminiscent of an approach from the 1980s, seemingly ignoring calls for more systemic, adaptive, and integrated approaches to managing nature. Article 8, with its focus on pollinators, is an exception to this. Ecosystem-based approaches, nature-based solutions, and co-benefits of

restoration for other environmental and societal objectives are mentioned, but the text does not elaborate on their implementation. Despite this, the NRL holds considerable potential to operate at ecosystem levels, providing widespread societal benefits, particularly through the increased supply of ecosystem services (*10*). Enhancing landscape structure and rewetting peatlands can increase the resilience of agricultural ecosystems to droughts and pests, and restoring pollinator populations can have direct positive impacts on agricultural production. Similarly, reconnecting rivers with their floodplains can mitigate flood risks (*11*); increasing urban green spaces can benefit urban climate and people's health; increasing forest diversity can enhance resilience to extreme events; and restoring marine ecosystems can benefit recreation (*12*).

AVOIDING PITFALLS

A recurring problem with the implementation of European environmental legislation and policies is the gap between targets and effective implementation options. HD, BD, WFD, and MSFD have so far not achieved their aims, and neither has the BS (see table S2). Reasons are manifold. Besides shortcomings in aims and approaches (see table S3, a to g), a common denominator is the lack of resources needed to implement them successfully, including funding, human resources, appropriate planning procedures, and administrative capacities for implementation. The passing of legislation and policies have not always been followed by the provision of appropriate resources and capacity-building for implementation and monitoring. The NRL encounters similar challenges because it is even more ambitious. Implementation at the national level must therefore assure a stringent procedure and a resilient funding structure, as suggested by the original Commission proposal. Although the targets are legally binding, the measures to achieve them will be voluntary actions by land and water owners and managers, who would need to accept co-responsibility and possess the capacity to respond. This requires not only financial investments but also supportive institutions for cooperation, peer-to-peer learning, business models that support land-use change, and societal acceptance to work with nature.

The required resources are not exclusively of public origin. After the NRL's approval, the EU and member states are tasked with mobilizing private financing of restoration, endorsing suitable business models that incorporate cost recovery (*13*). These may involve refined carbon credit trading, collaboration with insurance companies to mitigate flood or drought risks, or customized options for investing in nature. The European Investment Bank, and its enhanced capacity to offer advisory services alongside conventional financing, could assume a more prominent role in this regard.

It will be of equal importance to acquire public funds for restoration of nature from other components of the EU budget—in particular, regional development and agriculture. So far, despite the installment of relevant instruments, the CAP has not succeeded in achieving the aims of HD, BD, and WFD. The CAP is unlikely to contribute sufficiently to the NRL implementation if its support schemes are not modified to strengthen the ambition of measures, strictly enforce cross-compliance, and increase funding for focused measures. A specific clause is granting member states full flexibility in using or foregoing CAP or CFP funds for NRL implementation. Using these funds could potentially offer unprecedented, cost-efficient opportunities for both the NRL and the CAP and CPF. The CAP's agriculture-environment-climate measures, along with the somewhat less ambitious "Eco-schemes," could support habitat restoration and the recovery of pollinator populations. Implementing the NRL in farmlands is also vital for achieving various goals, including river-floodplain connectivity, river to coast-marine connectivity (through controlled floods), peatland targets (through alternative agricultural schemes such as paludiculture), and even urban restoration (by maintaining urban and peri-urban green and blue spaces). Simultaneously, addressing climate change in agriculture necessitates restoration measures such as landscape water storage, reduced livestock densities, and diminished nitrogen inputs.

"Although the targets are legally binding, the measures to achieve them will be voluntary actions..."

The trilogue negotiations have introduced two further elements that substantially weaken the NRL. First, member states may permanently deprioritize restoration actions in areas used for other targets such as renewable energy infrastructure and military facilities. Second, the inclusion of an "emergency brake" enables member states to temporarily suspend NRL implementation in farmland, over their entire area, under exceptional circumstances that affect land availability for agricultural production. However, an evaluation of the NRL planned for 2033 could result in legislative proposals for amendments, including a better coherence with other legislation or policies.

Translating ambitions into actions still requires a close alignment with both existing and emerging European legislation and policies. Stability in the legislative developments is crucial, considering that nature restoration requires long-term perspectives. Provision of funding schemes will determine whether the NRL will address current pressures and drive much-needed transitions. Given the urgency of global crises, Europe cannot afford to delay; the opportunity to install and implement an ambitious law, and the opportunity to show global leadership, should not be missed. ■

REFERENCES AND NOTES

1. European Commission, "Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee: The state of nature in the European Union, Report on the status and trends in 2013 - 2018 of species and habitat types protected by the Birds and Habitats Directives," COM(2020) 635 final (Publications Office of the European Union, 2020).
2. G. Pe'er *et al*., "Scientists support the EU's Green Deal and reject the unjustified argumentation against the Sustainable Use Regulation and the Nature Restoration Law" (2023); https://tinyurl.com/4ttp8mds.
3. R. Fischer et al., *Land Use Policy* **127**, 106546 (2023).
4. M. Nilsson *et al*., *Environ. Policy Gov.* **22**, 395 (2012).
5. S. Rigal *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2216573120 (2023).
6. N. Pettorelli *et al*., *J. Appl. Ecol.* **58**, 2384 (2021).
7. S. Vallecillo *et al*., *"EU-wide methodology to map and assess ecosystem condition: Towards a common approach consistent with a global statistical standard," EUR 31226 EN* (Publications Office of the European Union, 2022).
8. P. Borrelli, P. Panagos, D. Wuepper, *Int. Soil Water Conserv. Res.* **8**, 102 (2020).
9. S. Larsen *et al*., *PLOS ONE* **14**, e0213227 (2019).
10. R. S. DE Groot *et al*., *Conserv. Biol.* **27**, 1286 (2013).
11. J. J. Opperman *et al*., *Science* **326**, 1487 (2009).
12. S. Pouso, S. Ferrini, R. K. Turner, M. C. Uyarra, Á. Borja, *Front. Mar. Sci.* **5**, 2018.00375 (2018).
13. C. M. Duarte *et al*., *Nature* **580**, 39 (2020).

ACKNOWLEDGMENTS

This work was supported by the European Commission's Horizon 2020 programme, grant agreement 101036337 (MERLIN); European Commission's Horizon 2020 programme, grant agreement 101036484 (WaterLANDS); European Commission's Horizon 2020 programme, grant agreement 101036849 (SUPERB); European Commission's Horizon 2020 programme, grant agreement 101037097 (REST-COAST); European Commission's Horizon Europe programme, grant agreement 101059877 (GES4SEAS); European Commission's Horizon Europe programme, grant agreement 101060816 (Agroecology-TRANSECT); and the German Research Foundation (DFG-FZT 118, 202548816).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk1658

10.1126/science.adk1658

CONSCIOUSNESS

Chasing an equation for awareness

A writer seeks connections between consciousness and fundamental physics

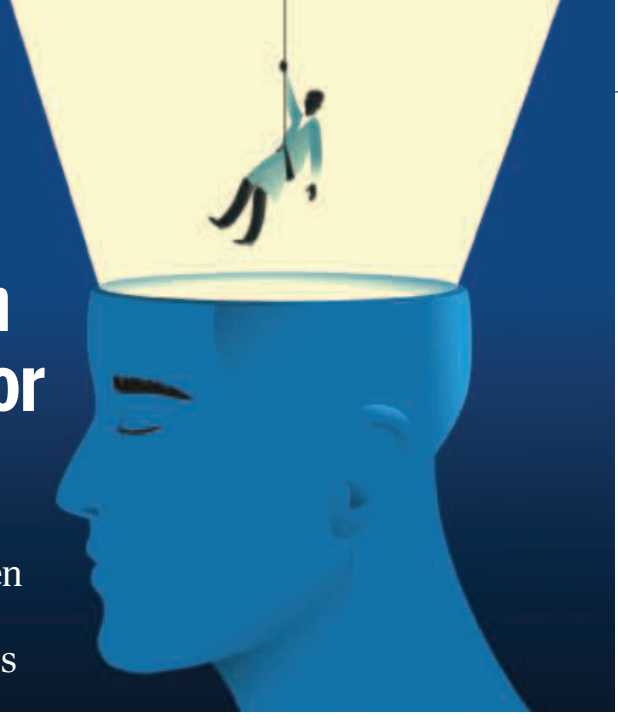
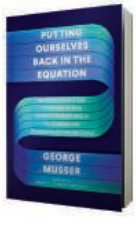


ILLUSTRATION: MOOR STUDIO

By **Chaz Firestone** and **Ian Phillips**

Science begins with mystery. What causes lightning? How did this mold stop bacterial growth? Why do we age? Arguably, the two greatest mysteries are the cosmos and consciousness—the vast world out there and the vibrant world within. Scientists captivated by one can be called to study the other, seduced by the thought that these mysteries are connected. Science writer George Musser's book *Putting Ourselves Back in the Equation* reviews their progress: Can physics unlock the mystery of consciousness? Does consciousness underlie fundamental physics?



Putting Ourselves Back in the Equation
George Musser
Farrar, Straus and Giroux, 2023. 336 pp.

The result is an ambitious but ultimately disappointing tour, peppered with breathless encounters with well-known scientists. Representative of the cast is MIT's Max Tegmark, who tells Musser: "If you look at the problems that we're still stumped on in foundational physics, pretty much all of them trace back to consciousness."

One puzzle is the quantum measurement problem. The basic formalism of quantum mechanics describes a world of superpositions: combinations of states represented by a wave function. Consider Schrödinger's celebrated cat. The relevant equations pronounce it both dead and alive; but when we open its box, it looks either dead or alive, never both. Multiplicity collapses into singularity—how? A radical solution is that consciousness is the culprit. "The mind reaches out," writes Musser, "grabs particles that are poised between possibilities, and tells them, Choose!"

Other solutions avoid such mental magic. "Many-worlds" interpretations—favored by many contemporary physicists, including Tegmark himself—eschew collapse entirely, eliminating the mind's putative role. Musser charges these views with incoherence (even comparing them to conspiracy theories) but does too little to substantiate his case.

The book's primary theme is how physicists are contributing to understanding the mind, continuing "a long history of physicists colonizing other fields." The metaphor is apt. It recalls an episode from Barbara Kingsolver's *The Poisonwood Bible*, where an American colonial missionary ignores the local horticultural expertise of his Congolese housekeeper. The upshot? A nasty rash from poisonwood sap and a flooded garden.

Musser integrates physics with neuroscience, economics, botany, chemistry, philosophy, mathematics, ornithology, and more. Yet a key source of local expertise is notably absent—psychology, the science of the mind. An early example: Musser rightly applauds physicists' contributions to artificial neural networks but is overly credulous of their implications, declaring that "ChatGPT and DALL-E are already able to do things that seem to be coming from deeply felt experience" and "are starting to demonstrate a generalized intelligence like that of humans." Interviewing more psychologists—experts in both feelings and intelligence—could have exposed the considerable gulf remaining (*1*, *2*).

What about consciousness itself? Why some states associated with felt experience (the pain of a headache, the sight of a sunset) and others not? Musser's focus is integrated information theory, or IIT. IIT begins with five allegedly self-evident "axioms" of awareness: consciousness exists, and it is structured, specific, unified, and definite. It then derives "postulates" concerning the causal structure of conscious systems, identifying consciousness with integrated information—information that cannot be reduced to that in a system's parts. Finally, IIT offers a mathematical measure of this quantity, Φ: an equation for awareness.

However, despite its enthusiasts, IIT has profound problems. Its axiomatic basis is specious (those that are not trivial are not self-evident), and grave doubts surround its testability and Φ's definability (*3*–*5*). Musser belatedly mentions Scott Aaronson's influential critique—that IIT implausibly assigns vast amounts of consciousness to DVD players and inactive logic gates—but seems unconcerned by this (devastating) result.

Of course, a theory of consciousness must do more than distinguish humans from DVD players; it must detail when, why, and to what degree we are conscious. Here, psychology's absence is most glaring. Over the last century, psychological research has revealed innumerable phenomena of consciousness, from paradigms that alter awareness (attentional blink, inattentional blindness) to methods that render stimuli unconscious (masking, flash suppression); from extraordinary disorders of consciousness, such as blindsight or neglect, to rigorous studies of metacognition ("awareness of awareness").

Although controversial, such phenomena are the data that any scientific theory of consciousness must account for. That is why all serious theories say something about them, including global neuronal workspace theory, higher-order theories, and recurrent theories. Yet these phenomena and ideas are almost nowhere in the book. Of course, psychology has not solved consciousness; but one cannot hope to unravel awareness without confronting these data.

Consciousness is genuinely mysterious. So is fundamental physics. But hoping that physics can solve consciousness while excluding other approaches is only a recipe for more mystery, not less. ■

The reviewers are at the Department of Psychological and Brain Sciences and the Department of Philosophy, Johns Hopkins University, Baltimore, MD 21218, USA.
Email: chaz@jhu.edu; ianbphillips@jhu.edu

REFERENCES AND NOTES

1. B. M. Lake, T. D. Ullman, J. B. Tenenbaum, S. J. Gershman, *Behav. Brain Sci.* **40**, e253 (2017).
2. E. Yiu, E. Kosoy, A. Gopnik, *Perspect. Psychol. Sci.* 10.1177/17456916231201401 (2023).
3. T. Bayne, *Neurosci. Conscious.* **2018**, niy007 (2018).
4. J. R. Hanson, S. I. Walker, *Neurosci. Conscious.* **2023**, niad014 (2023).
5. M. Michel, H. Lau, *PhiMiSci* **1**, 8 (2020).

10.1126/science.adl3136

ANTHROPOLOGY

Object lessons

An anthropologist confronts the history and ubiquity of human-made items

By **John Robb**

There are lots of ways to tell any story, and the more important a story is, the more ways it can be told. How objects define humanity and what they reveal about big patterns in human history—the topics under discussion in anthropologist Chip Colwell's eminently readable book *So Much Stuff*—is one such story. These types of tales require an author to consider long-term trends in technology, skill, and art and are often naively framed as resulting from some inherent force of progress driven by human ingenuity. Authors like Colwell, who approach material culture from social science perspectives such as anthropology, archaeology, and the history of technology, deconstruct these frameworks and seek to show the roles played by social contexts and networks.

Such stories must also address long-term trends in the quantity of objects humankind has owned over time. Why, for example, does the amount of things humans have keep increasing?

When you put both parts of this narrative together, there are two basic ways to spin it: as a utopian story of progress—of technology and capitalism's ever-expanding ability to furnish our every need and want—or as a pessimistic narrative in which the things we create and own suffocate us under their superabundance. In *So Much Stuff*, Colwell flirts with both angles.

The book frames the story of material culture as a series of leaps ahead, each of which builds on previous ones to move people from simple beginnings toward civilization, sophistication, and modernity. This is a familiar story of social evolution, drawing not only on cultural anthropology and archaeology but also on animal behavior, psychology, and cognitive research. Colwell tells it well in a chatty style. There is clearly a fair amount of academic thought underpinning the book, but it is worn lightly, with a greater emphasis on personal storytelling than on oppressing the reader with learned apparatus.

The book features many of the irresistible greatest hits of human material culture. These include Acheulean hand axes, the Blombos Cave engraved ochre, Ice Age cave art, Göbekli Tepe, Ötzi the Iceman, the Northwest Coast potlatch, the Industrial Revolution, colonialism, and resistance. It is framed in terms of a handful of qualitative milestones: making tools, giving things meanings, and the accelerating accumulation of things.

One of Colwell's premises is that all our things are problematic in some way. A running theme is that objects enable humans to do things and to be human while also confining or imprisoning us. This does not happen through any one object alone, Colwell argues, but through the cumulative mass of all of them. Together, our things strip the planet of resources, cost energy to use and maintain, generate enormous quantities of waste—and, potentially, reorient our senses and reactions in ways we might not like and that drive some people to seek out minimalist lifestyles.

Although present throughout the book, our dysfunctional relationship with things is never fully explored. Colwell can hardly be blamed for this; the overwhelming bulk of the anthropological literature on material culture is organized around case studies of particular objects, with almost none scrutinizing the effects of material culture as a total system.

The book also skates over theoretical dilemmas relatively lightly, usually by invoking universalizing generalizations. For instance, anthropological theorists might quibble about whether humans have an innate desire to accumulate things, and economists might wonder about blaming overabundance on the advertising industry rather than on the underlying economic system generating the need to create consumers. But stories such as how advertising developed tactics for stimulating consumer desire are fascinating and succeed in posing some thought-provoking questions. Why, for example, does religion have to have a material medium?

The book ends on one of its strongest points: the honest, personal account of Colwell's attempt to streamline his material life and get rid of things and how the system of things we live in defeated it. This account does not solve the dilemma the book poses, but it highlights it in ways that will resonate with most readers. Ultimately, *So Much Stuff* provides an engrossing introduction for nonexperts into the big questions of material culture studies. ■

The things we make, acquire, and accumulate reveal much about human civilization.

The reviewer is at the Department of Archaeology, University of Cambridge, Cambridge CB2 3DZ, UK. Email: jer39@cam.ac.uk

So Much Stuff: How Humans Discovered Tools, Invented Meaning, and Made More of Everything
Chip Colwell
University of Chicago Press, 2023. 304 pp.

10.1126/science.adl1546

LETTERS

Brazil's state of Amazonas has suffered from an extreme, months-long drought.

PHOTO: AFP VIA GETTY IMAGES

Edited by **Jennifer Sills**

Extreme drought threatens the Amazon

The Amazon River basin is the Brazil's largest source of freshwater (*1*). Traditional and local communities rely on the Amazon and its tributaries (*2*) for basic resources such as water and food, as well as for transportation that facilitates the delivery of medical care and access to education (*2*–*5*). However, since June (*6*), the Amazon region has suffered an extreme drought, which has led some sections of the river system to dry up completely (*6*–*8*). To meet global sustainability goals, Brazil must provide immediate relief to communities affected by the drought as well as implement comprehensive policies and practices to protect the Amazon basin from extreme droughts.

The drought in the state of Amazonas, the most intense in the last 120 years, has been exacerbated by other events (*6*, *7*). El Niño and unusually warm waters in the Atlantic have contributed to increased global temperatures, which have intensified the dry season in the Amazon (*2*, *9*). Deforestation resulting from human activities such as mining, agricultural expansion, and extensive livestock farming has also burdened the regional climate balance (*9*).

Given that the drought could persist until 2024 (*9, 10*), the Brazilian state of Amazonas has declared a state of emergency (*2*, *9*, *11*). The government has distributed basic supplies and financial support to rural producers who cannot hunt, fish, or earn money by trading goods sourced from the rivers (*11*). However, it is essential to go beyond short-term aid and address the problem's causes. Long-term strategies should include integrating economic policies in the Amazon with Brazil's climate goals, revitalizing degraded areas, and combating socioeconomic inequality. Additionally, Brazil and the global community must promote international collaboration in the fight against climate change.

The effects of the drought undermine compliance with Sustainable Development Goals 2, 3, 4, 6, 13, and 15 (*12*), which provide targets for reducing hunger, maintaining health, supporting education, assuring access to clean water, taking action to address climate change, and protecting ecosystems, respectively. Brazil should take urgent action to manage, mitigate, and prevent extreme droughts in the Amazon region.

Felipe P. Ottoni[1,2,3,4]*, Carlos Thiago S. Filgueira[1], Brenda N. Lima[1], Lucas O. Vieira[1,3], Filipe Rangel-Pereira[5], Rafael F. Oliveira[1,4]

[1]Laboratório de Sistemática e Ecologia de Organismos Aquáticos, Centro de Ciências de Chapadinha, Universidade Federal do Maranhão, Chapadinha, MA, Brazil. [2]Programa de Pós-Graduação em Ciências Ambientais, Universidade Federal do Maranhão, Chapadinha, MA, Brazil. [3]Programa de Pós Graduação em Biodiversidade e Biotecnologia da Amazônia Legal, Universidade Federal do Maranhão, São Luís, MA, Brazil. [4]Programa de Pós-Graduação em Biodiversidade e Conservação, Universidade Federal do Maranhão, São Luis, MA, Brazil. [5]Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil.
*Corresponding author. Email: fpottoni@gmail.com

REFERENCES AND NOTES

1. R. E. Reis *et al*., *J. Fish Biol*. **89**, 12 (2016).
2. A. Schröder, "Amazon drought cuts river traffic, leaves communities without water and supplies," *Mongabay* (2023).
3. M. E. Portela, "Seca no Amazonas impacta distribuição de alimentos e energia no Brasil," *Metrópoles* (2023) [in Portuguese].
4. L. Prazeres, "Não consigo mais plantar: como seca histórica na Amazônia afeta mulheres indígenas," *BBC News Brasil* (2023) [in Portuguese].
5. L. Vasconcelos, "Estiagem afeta o cotidiano de alunos do interior do Amazonas," *A Crítica* (2023) [in Portuguese].
6. V. Aguiar, "Seca no Amazonas: Rio Negro atinge menor nível em 121 anos, com apenas 13,59 metros," *CCN BRASIL* (2023) [in Portuguese].
7. F. Maisonnave, "Rivers in the Amazon fall to their lowest levels in 120 years," *The Sydney Morning Herald* (2023).
8. G. Campos, "Rio Solimões vira deserto em meio a seca histórica no Amazonas," *Jornal Opção* (2023) [in Portuguese].
9. F. Milhorance "Historic drought in the Amazon, explained," *Diálogo Chino* (2023).
10. E. Barros, "'Without water, there is no life': Drought in Brazil's Amazon is sharpening fears for the future," *CLIMATE* (2023).
11. A. Góes, "Federal task force to bring humanitarian aid to drought-affected towns in AM," *Cenarium Amazon Magazine* (2023).
12. United Nations, "The 17 goals" (2015); https://sdgs.un.org/goals.

10.1126/science.adm8147

EU policies undermine environmental targets

In July, the United Nations announced that progress toward achieving the global Sustainable Development Goals by 2030 is falling substantially short (*1*). Countries around the world have also failed to meet the Aichi Biodiversity Targets (*2*). Climate anomalies are breaking records, and the global warming target of 1.5°C set in the Paris Climate Agreement is likely to be surpassed soon (*3*). Despite these shortfalls, the EU and its Member States are pursuing legislation that undermines global goals.

The EU Commission recently proposed a nature restoration law, which would meet the international commitment under the UN Global Biodiversity Framework to restore 30% of degraded ecosystems in certain landscapes by 2030, but after a negotiation with the EU Council and the Parliament, the proposed targets have been substantially weakened (*4*). The EU co-legislators agreed to restore at least 20% of land and 20% of marine areas by 2030, well short of the 30% goal, and the corresponding restoration may focus on already existing European protected areas of the Natura 2000 system. In agroecosystems, the goal is merely to show a "positive

trend" for selected indicators by 2030, and a loophole would allow an emergency brake—i.e., a suspension of efforts—if ecosystem objectives threaten the agricultural production needed for EU food consumption. In addition to making the restoration law less ambitious, the EU Parliament rejected draft legislation that would have halved the use of pesticides by 2030 (*5*).

EU Member States are also hindering progress toward global environmental goals. In Germany, Chancellor Olaf Scholz and the state premiers have adopted the Germany Pact, which will speed up the implementation of construction projects and reduce time-consuming environmental impact assessments and participation procedures (*6*). According to Germany's largest conservation nongovernmental organization, NABU, the pact threatens key environmental gains made in recent decades (*7*). Meanwhile, the EU Commission's infringement reports show that many Member States regularly fail to implement existing environmental policies. For example, Belgium has failed to bring its national legislation into line with the EU's Environmental Impact Assessment Directive (*8*).

As a major contributor to the global environmental crisis, the EU should take responsibility for its contributions and lead efforts to mitigate the damage. Member States, especially large economies such as Germany, should work to show that human well-being can be reconciled with a substantial reduction in the ecological footprint, both nationally and globally. Rich countries must demonstrate that international environmental agreements are true commitments and not just platitudes.

Pierre L. Ibisch
Centre for Econics and Ecosystem Management, Eberswalde University for Sustainable Development, Eberswalde, Germany.
Email: pierre.ibisch@hnee.de

REFERENCES AND NOTES

1. United Nations, "The Sustainable Development Goals report 2023: Special edition" (2023); https://unstats.un.org/sdgs/report/2023/.
2. J. E. Hansen *et al.*, *Oxford Open Clim. Change* **3**, kgad008 (2023).
3. Secretariat of the Convention on Biological Diversity, "Global Biodiversity Outlook 5: Summary for policymakers" (Montreal, 2020); https://www.cbd.int/gbo/gbo5/publication/gbo-5-spm-en.pdf.
4. European Parliament, "EU Nature restoration law: MEPs strike deal to restore 20% of EU's land and sea" (2023); https://www.europarl.europa.eu/news/en/press-room/20231031IPR08714/eu-nature-restoration-law-meps-strike-deal-to-restore-20-of-eu-s-land-and-sea.
5. European Parliament, "No majority in Parliament for legislation to curb use of pesticides" (2023); https://www.europarl.europa.eu/news/en/press-room/20231117IPR12215/no-majority-in-parliament-for-legislation-to-curb-use-of-pesticides.
6. Advant Beiten, "Pakt für Planungs-, Genehmigungs- und Umsetzungsbeschleunigung zwischen Bund und Ländern" (2023); https://www.lexology.com/library/detail.aspx?g=5a321964-1916-44dd-acf8-cd4afae7543a [in German].
7. NABU, "Scharfe Kritik an Pakt für Planungsbeschleunigung: Beschlüsse öffnen Tür für neue Ära der Naturzerstörung" (2023); https://www.nabu.de/news/2023/10/34054.html [in German].
8. European Commission, "June infringement package: key decisions" (2023); https://ec.europa.eu/commission/presscorner/detail/%20en/inf_23_2707.

10.1126/science.adm8856

Third Pole needs more than legal protection

The Qinghai-Tibet Plateau (QTP), known as Earth's Third Pole and the Asian Water Tower, is the source of nine major Asian rivers, supporting 1.5 billion people downstream with freshwater, food, and essential ecosystem services. The region plays a pivotal role in global climate regulation and biodiversity preservation (*1*). Because of its high altitude, low temperatures, and arid conditions, the QTP's ecosystems are inherently fragile and exceptionally responsive to human activities. To better safeguard the QTP within China, the Chinese government has enacted the Ecological Protection Law, effective since September (*2*). However, comprehensive protection of the QTP requires changes and cooperation beyond this legislation.

Between 1980 and 2020, the QTP population within China grew to 156% of its original size, urban construction nearly tripled, roads quadrupled, and tourist visits surged, reaching 70 million in 2019—five times the local population (*3*). This heightened human activity poses substantial challenges to the QTP's ecology. Glaciers have retreated, severe land desertification has occurred, 70% of grasslands have been degraded, biodiversity threats have increased, and plastic waste has proliferated (*4*, *5*).

The Ecological Protection Law emphasizes holistic protection of the plateau's ecosystems and wildlife resources (*2*). The legislation underscores the need to intensify surveys and monitoring of natural resources and ecological conditions. The law also mandates enhanced biodiversity conservation in pivotal areas such as the Three Rivers Source, Qilian Mountains, and Mount Qomolangma, using measures such as establishing national parks and nature reserves. To facilitate implementation, the text precisely delineates the roles of government departments and proposes the formulation of ecological zoning and control plans, alongside a list of industries prohibited from operating in the QTP based on the environmental risks they pose.

Although the law is a step forward, the ecological protection of the QTP still lacks sufficient green industries, eco-friendly tourism practices, sustainable technology, multi-stakeholder engagement, and international collaboration. China should foster green industries such as ecotourism, ecological agriculture, clean energy, and cross-border trade. To address the substantial solid waste produced by tourists and outdoor enthusiasts, it is imperative to intensify environmental education and legal awareness among visitors as well as to augment waste recycling infrastructure. The Chinese government should provide increased science and technology funding to develop biodiversity-friendly and climate-adaptive technologies for high-altitude mountain living. To facilitate the monitoring now required by law, local authorities should collaborate with scientists and environmental organizations to formulate region-specific law enforcement guidelines and conduct regular evaluations of ecological conditions and the efficacy of protection and restoration efforts. Given the intricate topography, frequent geological hazards, and legal disparities across multiple nations in the QTP and surrounding mountainous regions, strengthened collaboration among neighboring countries is essential in scientific research, biodiversity preservation, ecological compensation, and international river governance.

Haimeng Liu[1]* and Hao Chen[2]
[1]Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China. [2]Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China.
*Corresponding author. Email: liuhm@igsnrr.ac.cn

REFERENCES AND NOTES

1. T. Yao *et al.*, *Environ. Dev.* **3**, 52 (2012).
2. "Qinghai-Tibet Plateau Ecological Protection Law of the People's Republic of China" (2023); http://faolex.fao.org/docs/pdf/chn217026.pdf [in Chinese].
3. H. Liu *et al.*, *Earth's Future* **11**, e2022EF003452 (2023).
4. C. Li *et al.*, *Land Degrad. Dev.* **32**, 2371 (2021).
5. Q. Li *et al.*, *Environ. Int.* **175**, 107956 (2023).

10.1126/science.adl5035

ERRATA

Erratum for the Research Article "Agriculture and hot temperatures interactively erode the nest success of habitat generalist birds across the United States" by K. S. Lauck *et al.*, *Science* **382,** eadn2847 (2023).

Published online 15 December 2023
10.1126/science.adn2847

Erratum for the Research Article "Ternary NiMo-Bi liquid alloy catalyst for efficient hydrogen production from methane pyrolysis" by L. Chen *et al.*, *Science* **382,** eadm9972 (2023).

Published online 8 December 2023
10.1126/science.adm9972

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

The location of i poaching of threate African pangolins has shifted over the past decade.

PHOTO: MATTHIJS KUIJPERS/ALAMY STOCK PHOTO

CONSERVATION

Pinpointing pangolin poaching

Over the past decade, Asian pangolin populations have declined due to demand from unsubstantiated use in traditional medicine; at the same time, poaching of African pangolins has increased rapidly. Identifying the source of poached pangolins and the path of transit of pangolin scales could greatly improve our ability to reduce illegal trade of these animals, which at current rates could lead to their extinction. Tinsman *et al*. used a genomic approach to identify the populations of origin for white-bellied pangolin (*Phataginus tricuspis*) samples and seizures. They found a shift from West to Central Africa over the past decade and further identified Nigeria as an important hub. —SNV

Science, adi5066, this issue p. 1282

MESOSCOPIC PHYSICS

Propagating quantum information

Qubits, the building blocks of quantum computers, are usually implemented in stationary entities such as trapped ions or quantum dots. The job of propagating quantum information is usually given to photons. However, photons do not interact with each other, and it would be advantageous to implement a propagating ("flying") qubit using electrons. Assouline *et al*. achieved this goal by preparing and controlling single-electron flying qubit states in monolayer graphene. —JS

Science, adf9887, this issue p. 1260

THERMAL TRANSPORT

Directing the heat

Controlling thermal transport by switching devices is challenging because rapidly changing a material's thermal properties is difficult. Liu *et al*. demonstrate that a thermal switch can be made from lead zirconium oxide controlled with a small voltage. The material can be changed from ferroelectric to antiferroelectric with an electric field that also dramatically changes the thermal properties. This switch changes the thermal transport by a factor of two in one direction, creating a useful thermal switch. —BG

Science, adj9669, this issue p. 1265

MACHINE LEARNING

Getting rid of backpropagation

The recent development of large-scale deep neural networks (NNs) and other artificial intelligence (AI) applications is accompanied by growing concerns about the energy consumption needed to train and operate them. Physical NNs could become a solution to this problem, but the direct hardware implementation of conventional algorithms faces multiple difficulties. For instance, training NNs using conventional backpropagation algorithms is associated with challenges such as lack of scalability, complexity of operation during training, and dependency on digitally trained models. Inspired by the forward-forward algorithm, Momeni *et al*. report the practical demonstration of backpropagation-free training of wave-based physical NNs. Their work is an important

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

NEURODEGENERATION

Alzheimer's disease immunotherapies

Three monoclonal antibodies that target amyloid-β deposits in the brain have been found to slow cognitive decline in individuals with early Alzheimer's disease (AD). The antibodies, aducanumab, lecanemab, and donanemab, are thus the first disease-modifying treatments for AD. In a Perspective, Golde and Levey discuss the implications of this success, as well as the limitations of these treatments and the future work required to understand more about their mechanism of action. The authors highlight outstanding questions, including whether the clinical effects observed in the clinical trials, which finished after 18 months, will be maintained with continued treatment and if it will eventually be possible to predict and prevent adverse events such as amyloid-related imaging abnormalities. —SAL

Science, adj9255, this issue p. 1242

DEVELOPMENT

A multifaceted pioneer factor

Pluripotency transiently appears in early development and is controlled by pioneer transcription factors (TFs) such as OCT4 and SOX2. However, how these master TFs drive pluripotency progression in early embryos remains poorly understood. Li *et al*. investigated spatiotemporal chromatin occupancy of SOX2 in mouse early embryos across 4 days. In early blastocysts, SOX2 regulates the pluripotency program not just by opening global enhancers, but also by occupying preaccessible enhancers opened by early-stage-expressing TFs. SOX2 then redistributes and opens enhancers or primes enhancers for future activation when cells acquire naive and formative pluripotency. These data revealed flexible pioneer TF-chromatin interactions and a transitionary "prepluripotency" state that connects totipotency and pluripotency. —DJ

Science, adi5516, this issue p. 1258

GLOBAL WARMING

Coral reefs in peril

The past year has seen record-breaking sea surface temperatures, and extraordinary levels of heat stress have been recorded, particularly at coral reefs in the eastern tropical Pacific and wider Caribbean. In a Perspective, Hoegh-Guldeberg *et al.* discuss how marine heat waves and attendant heat stress on corals is likely to cause widespread coral bleaching and mortality. Unless global warming can be curtailed through reduced greenhouse gas emissions, coral reefs are likely to be severely affected. As much as 25% of ocean biodiversity could be affected if coral reef ecosystems collapse. The authors discuss approaches to better monitor the health of coral reefs and what management is needed to safeguard coral reef sites across the globe. —GKA

Science, adk4532, this issue p. 1238

MICROBIOTA

Never working alone

The microbiota plays an important part in host defense by excluding pathogenic species in a process called colonization resistance. This property is elusive and not endowed by one or two species alone. Spragge *et al*. discovered that colonization resistance is a higher-order effect of a diverse community of bacteria underpinned by essential key species such as *Escherichia coli* (see the Perspective by Radlinski and Bäumler). In vitro and in vivo experiments showed that given the right composition, a diverse microbiota will collectively consume the nutrients that an incoming species requires to grow and establish in a host. Colonization resistance is predictable because if the symbiont community encodes many of the same (or similar) proteins as the pathogen, then it provides better colonization resistance and can potentially deliver health benefits to the host. —CA

Science, adj3502, this issue p. 1259; see also adl5891, p. 1244

ANCIENT DNA

Reclaiming the Coast Salish woolly dog

Multiple Indigenous groups, such as the Coast Salish in the Pacific Northwest, bred and maintained dogs with a distinctive woolly undercoat, and hair from these dogs was used in weaving. Lin *et al*. sequenced the only known specimen of a Coast Salish woolly dog named Mutton from 1859 (see the Perspective by Orlando). Mutton was found to have limited introgression from European colonial dogs and otherwise lacked much genetic diversity, suggesting that these dogs were carefully reproductively managed to maintain their coats. By incorporating traditional knowledge, historical records, and genetic data, the cultural importance of these dogs and the role of colonialism in their loss are illuminated. This study joins the growing number attempting to incorporate Indigenous groups in historical studies. —CNS

Science, adi6549, this issue p. 1303 see also adm6959, p. 1236

VIRAL PALEOGENOMICS

Ancient DNA illuminates viral evolution

Marek's disease virus (MDV) causes tumors in chickens and has increased in virulence over the past century. Fiddaman *et al*. studied ancient DNA from chickens to reveal the evolutionary history of this once mild disease (see the Perspective by Duchene). They sequenced MDV DNA from 15 ancient chickens, finding that these strains were basal to modern ones and identifying multiple fixed genetic changes. Testing the ancient version of the gene *Meq*, a regulator of tumor formation, the authors found that virulence was greatly reduced compared with modern versions. These results shed light on the evolution of virulence in a major disease in modern agriculture. —CNS

Science, adg2238, this issue p. 1276; see also adl6094, p. 1245

STELLAR ASTROPHYSICS

A population of stripped helium stars

Removing the hydrogen-rich layers from a main sequence star exposes the helium-rich core. Such stripped helium stars are known at high and low masses, but not at intermediate masses, despite theoretical predictions that they should be common. Drout *et al*. used ultraviolet photometry to identify candidate stripped helium stars in two nearby dwarf galaxies and then observed 25 such candidate stars with optical spectroscopy (see the Perspective by Sundqvist and Sana). They found eight stars with spectra matching predictions for intermediate-mass stripped helium stars, and another eight with similar spectra that were blended with a main sequence companion. Most of these systems were shown to be binaries, in which the companion stars probably stripped the outer hydrogen-rich layers off the helium stars. —KTS

Science, ade4970, this issue p. 1287; see also adl5676, p. 1240

THERMOREGULATION

Consistent comfort

Clothing often helps to regulate comfort, but it is normally focused on keeping a person either warmer or cooler. Wang *et al*. developed a thermoregulatory

clothing system that combines an organic photovoltaic with bidirectional electrocaloric devices that are capable of heating or cooling (see the Perspective by Huang and Li). Both components are flexible, which is important for personal thermal regulation applications. The device runs off of sunlight, so no additional power sources are needed, and it could be useful in a variety of harsh environments. —BG

Science, adj3654, this issue p. 1291; see also adl5650, p. 1247

ECOLOGICAL SCIENCE

Shifting subsidies restructure food webs

Subsidies from an ecosystem's surrounding environments often play critical roles in structuring biological communities. Lin *et al.* show that on Orchid Island, which lies off the coast of Taiwan, a huge restructuring of its herpetofauna followed loss of an important marine-derived subsidy: green sea turtle eggs. Beach loss attributable to sea-level rise, habitat degradation, and predator-exclusion fences constructed to protect sea turtle nests caused predatory snakes that had been feeding on sea turtle eggs to shift inland to feed on the eggs of forest lizards. This study draws attention to the complex, cascading consequences of environmental pressures and single-species conservation efforts. —SN

Sci. Adv. (2023) 10.1126/sciadv.adj7052

IMMUNOLOGY

How interferon signaling goes viral

At the start of a viral infection, a few cells secrete type I interferons (IFNs), which stimulate transcription factors to promote antiviral protein accumulation. Korwek *et al*. measured how IFNs affect the signaling dynamics of transcription factors and antiviral proteins in cells treated with poly(I:C), a viral RNA analog. These data were then used to generate a model that explained how IFN prestimulation caused poly(I:C)–treated cells to switch from being IFN responding to IFN secreting, a critical conversion for mounting an effective antiviral response. —JFF

Sci. Signal. (2023) 10.1126/scisignal.abq1173

VACCINES

Weak T cell response in HIV vaccines

Vaccine-generated $CD8^+$ T cells could play an important role in containing HIV below detectable levels. In chronic infection, robust correlates of spontaneous immunologic control have been identified. Based upon preclinical work in SIV- or SHIV-challenged Rhesus macaques, $CD8^+$ T cells induced by vaccines or passive transfer of antibodies during acute infection appear to fully suppress or even eradicate lentiviral infection. Despite these advances, the correlates of this activity are not fully understood. Migueles *et al*. report that low antigen levels present on HIV-infected targets caused impaired degranulation and low $CD8^+$ T cell–mediated cytotoxicity. The TCR repertoire was polyclonal, and transduction of these TCRs conferred the same reduced functions. Thus, effective $CD8^+$ T cell responses in HIV/AIDS vaccination may require a strategy that drives further TCR clonal selection. —STS

Science, adg0514, this issue p. 1270

NK CELLS

Human memory NK cells

Natural killer (NK) cells are well known for their innate immune functions, but several studies have alluded to their ability to develop antigen-specific memory responses. Jost *et al.* characterized antigen-specific NK cell memory responses in humans exposed to HIV-1 or influenza virus. They determined that epitope specificity was mediated by activating CD94/NKG2C receptors engaging with peptides presented by the protein HLA-E. Antigen-specific memory NK cells also showed up-regulation of several biomarkers that help to define attributes of this subset. These findings provide critical insights into the mechanisms associated with antigen-specific memory NK cell responses that can potentially contribute to the development of vaccines and targeted therapies. —CNF

Sci. Immunol. (2023) 10.1126/sciimmunol.adi3974

step in optimizing the energy-intensive training step in NNs for more efficient solutions for modern AI systems. —YS

Science, adi8474, this issue p. 1297

PLANT OPTOGENETICS

Illuminating ion interactions

Protons (H^+) and calcium ions (Ca^{2+}) are often shuttled across cellular membranes and act as signals. In plants, protons accumulate highly in the extracellular space, whereas Ca^{2+} stores can be found in the endoplasmic reticulum and other cellular compartments. H^+ and Ca^{2+} concentrations in the cytoplasm are often closely correlated during signaling events. Huang *et al*. introduced a light-gated, protist-derived ion channel into plant cells, allowing selective transport of H^+ but not Ca^{2+}. They found that induction of H^+ transport into the cytoplasm led to subsequent Ca^{2+} release. This work establishes a causal relationship between intracellular pH and Ca^{2+} signaling in guard cells and offers a new tool for manipulating ion movement in other contexts. —MRS

Science, adj9696, this issue p. 1314

KIDNEY DISEASE

Multiomics for markers of maladaptation

Although acute kidney injury (AKI) commonly affects hospitalized patients and is associated with adverse outcomes, it has been understudied in humans. Wen *et al.* integrated single-nucleus RNA-sequencing data from human kidney biopsy specimens with plasma proteomic data to identify markers of both proximal tubule maladaptation and healthy states. The authors validated these markers in independent patient cohorts as well as in mouse models of ischemia reperfusion injury and aristolochic acid–induced nephropathy. Proximal tubule maladaptation appears to occur in response to injury from a number of causes, and the markers of proximal tubule health and disease identified here may be useful for both prognostic and drug development purposes. —MLN

Sci. Transl. Med. (2023) 10.1126/scitranslmed.ade7287

ATMOSPHERES

Supercharging new particle formation

How are new particles formed in the air above the oceans, where ammonia, an important species in the process, is not very abundant? He *et al*. report that iodine oxoacids, which are plentiful in marine environments, can substantially increase the rate of new particle formation in the low-ammonia conditions commonly found in pristine marine and polar regions. This effect could be particularly important in low-level marine stratocumulus clouds, which reflect a large fraction of incident solar radiation back into space and have an important influence on global radiation balance and climate. —HJS

Science, adh2526, this issue p. 1308

ARCHAEOLOGY

Animal husbandry on the Tibetan Plateau

Defining the timing of yak domestication and the appearance of taurine cattle on the Tibetan Plateau has been hampered by the scarcity of samples of either species found in secure archaeological contexts. Recent archaeological and genomic research conducted by Chen *et al*. on animal remains recovered from the Bangga site in central Tibet identified the presence of both domesticated yak and taurine cattle from 2500 years ago. The authors further discovered evidence for yak-cattle hybrids, and suggest that these findings offer evidence of husbandry strategies by ancient herders that improved the fitness of these species in this high-elevation environment. —MSA

Sci. Adv. (2023) 10.1126/sciadv.adi6857

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

CHILDCARE

Does it take a "village" to raise a child?

Dominant psychological attachment theories originate from educated, Western populations and center on the mother-child relationship rather than on a "village" structure. However, during most of our evolutionary history, humans lived in hunter-gatherer communities that supported each other while rearing infants and toddlers. Chaudhary *et al.* scrutinized the Mbendjele contemporary hunter-gatherer communities in the Republic of Congo. Mbendjele children regularly received attention to soothe crying, tactile care, and daily physical contact from various caregivers far beyond their biological parents. This suggests that our ancestors may have enjoyed a more extensive support system that alleviated the high caregiving pressures that parents in contemporary Western countries now endure alone. More research is needed to understand the psychological returns on larger care networks for contemporary, Western children. —EEU

Dev. Psychol. (2023) 10.1037/dev0001601

COVID-19

Thinking two steps ahead

T lymphocytes are immune cells that recognize pathogens and mount immune responses to infections such as those by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus that causes COVID-19. Tarke *et al.* investigated how T cells recognize similarities between viruses within the same family, in particular coronavirus strains that caused the COVID-19 and MERS pandemics, along with common cold coronaviruses. Certain viral fragments called epitopes are conserved among different coronaviruses. First, viral epitopes were identified that are recognized by T cells, followed by computational modeling to reveal the sequences shared between SARS-CoV-2 and other

Nitrate ions made by marine organisms may cause marine aerosol nucleation and enhance cloud formation over the ocean.

MARINE CLOUDS

Outlook cloudy

The possible effects of marine microorganisms on cloudiness above the ocean, and the feedback that such effects might in turn have on climate, have been investigated for years, centering mostly around emissions of dimethyl sulfide. Chamba *et al.* changed the focus to nitrate ions and their influence on marine aerosol nucleation. Their observations from experiments conducted in the South Pacific suggest that nitrate ions are key species involved in aerosol nucleation in the pristine marine atmosphere and that nitrate ion formation is related to short-term microbial processes and dimethyl sulfide formation. —HJS

Geophys. Res. Lett. (2023)
10.1073/pnas.2308696120

coronavirus strains. These findings may help in the design of vaccines that elicit broader protection against multiple coronaviruses and, by extension, help to safeguard against future pandemic threats. —PNK

Cell Rep. Med. (2023)
10.1016/j.xcrm.2023.101088

SCIENCE EDUCATION

Gateway not gatekeeper

General chemistry (GC) is known as a "gatekeeper" course that most STEM students take in their first semester of college, and this gatekeeper effect affects historically marginalized students the most. Edwards *et al.* examined students' persistence from GC1 to GC2 and found that GC1 performance alone (i.e., class grade) did not fully explain students' persistence to GC2. Instead, students' social belonging in GC1, referring to a student's feeling of social connectedness to peers, instructors, and course environment, was influential and differed across genders. These results confirm that there is more to CG success than a grade, and suggest that students' social belonging should be considered across the chemistry curriculum. —MMc

J. Chem. Educ. (2023)
10.1021/acs.jchemed.2c01048

REPRODUCTION

Maternal front-loading

White-browed sparrow weavers (*Plocepasser mahali*) build large, messy nests of twigs in acacia trees found in African savannahs. Each social group of these birds is dominated by a breeding pair, and offspring are cared for with the assistance of variable numbers of helpers. Capilla-Lasheras *et al.* have monitored breeding in this species over 10 years. During that time, the authors found that if more female helpers were available, then the breeding female laid larger eggs but was subsequently less involved in feeding young. More male availability for care did not result in greater egg investment because males are just not as good as females at feeding young. Somehow, mother sparrow weavers assess the value of postnatal help, thus enabling prenatal investment in the eggs. —CA

PLoS Biol. (2023)
10.1371/journal.pbio.3002356

NEUROSCIENCE

Cross-species transmission of emotions

Many animals take advantage of social cues to guide their behavior. By sensing such cues from conspecifics, they can detect threats before encountering them. Animals can also recognize the emotional expressions of other species. Such emotional eavesdropping often appears in interactions between humans and other animals, especially domesticated species. Kaźmierowska *et al.* found that interaction with a human caregiver who had recently undergone fear conditioning changed rats' behavior and activated their amygdala, the brain structure crucial for threat detection and orchestrating defensive responses. —PRS

Proc. Natl. Acad. Sci. U S A (2023)
10.1073/pnas.2302655120

INORGANIC CHEMISTRY

A weak but reactive iron carbene

The triple bond of metal complexes to carbene ligands is usually a low-spin singlet state, but Rao *et al.* now report a triplet carbene complex to iron. The oxidative addition of a low-coordinate iron(0) complex into the carbon–sulfur bond of a five-membered ring of thioazole-2-ylidene yielded a six-membered iron carbene ring. The long triple bond, along with spectroscopic and theoretical results, supports a triplet state. Unlike a typical singlet metal carbene, this complex can undergo thermal carbyne-coupling reactions that eliminate sulfur, as well as addition reactions with electrophiles such as carbon dioxide and diphenylsilane. —PDS

J. Am. Chem. Soc. (2023)
10.1021/jacs.3c09280

RESEARCH ARTICLE SUMMARY

DEVELOPMENT

Multifaceted SOX2-chromatin interaction underpins pluripotency progression in early embryos

Lijia Li†, Fangnong Lai†, Xiaoyu Hu†, Bofeng Liu†, Xukun Lu, Zili Lin, Ling Liu, Yunlong Xiang, Tristan Frum, Michael A. Halbisen, Fengling Chen, Qiang Fan, Amy Ralston, Wei Xie*

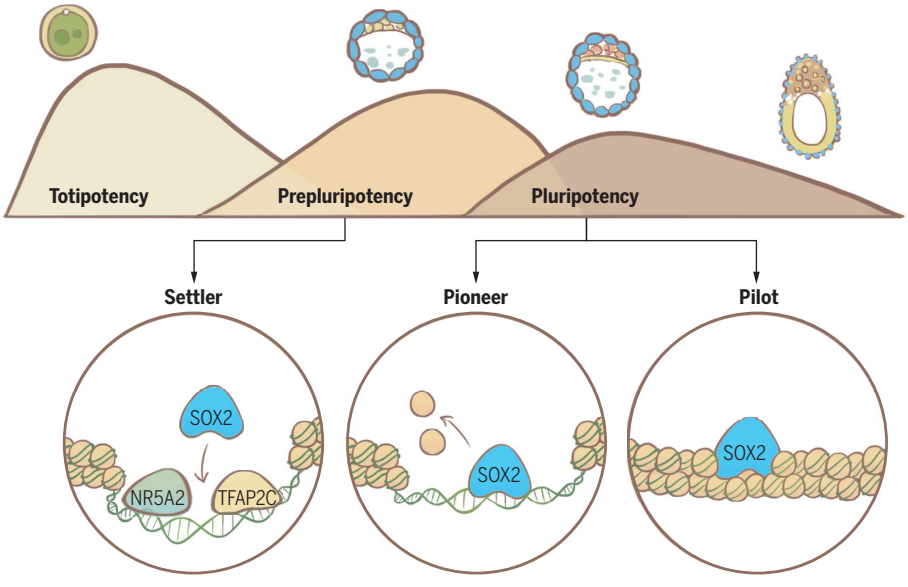
INTRODUCTION: During early mammalian development, a totipotent embryo undergoes the first cell fate decision to form a blastocyst that includes inner cell mass (ICM) and trophectoderm. ICM gives rise to epiblast, the origin of future embryonic lineages, and primitive endoderm. Trophectoderm subsequently differentiates to placenta. Pluripotency—the ability for a cell to give rise to all primary embryonic lineages—emerges within ICM and transits through several states: naive, formative, and primed pluripotency. Pioneer transcription factors (TFs), such as OCT4 and SOX2, which can bind and open closed chromatin, are crucial for pluripotency regulation. However, their regulatory circuitry is largely inferred from cultured cells. How the master TFs govern pluripotency progression in vivo remains challenging to study, largely because of the limited research materials from mammalian embryos.

RATIONALE: To study TF-gene interactions in early mammalian embryos, we applied CUT&RUN to capture the chromatin binding of SOX2, an early ICM marker, from embryonic day 3.5 (E3.5) to E7.5, which covers the entire progression of pluripotency in mouse embryos. Combined with RNA sequencing (RNA-seq) and assay for transposase-accessible chromatin sequencing (ATAC-seq) in *Sox2* knockout embryos and degron-tagged embryonic stem cells (ESCs), we investigated the roles of SOX2 in gene regulation and enhancer opening during pluripotency progression in vivo and compared this with in vitro conditions.

RESULTS: Our data revealed that SOX2 in E3.5 ICM has a regulatory circuitry distinct from that in all other pluripotent states and ESCs. Two massive relocalizations of SOX2-chromatin binding occurred when cells entered naive (E4.5 epiblast) and formative (E5.5 epiblast) pluripotent states, followed by a less dynamic transition to E7.5 ectoderm. These changes were accompanied by a critical role of SOX2 in regulating the ICM-trophectoderm transcription program and the following naive-to-formative pluripotency conversion.

Furthermore, we discovered that SOX2 manifests much more diverse binding modes at enhancers, which include "settler binding," "pioneer binding," and "pilot binding," beyond a simple "pioneer factor." SOX2 exhibits settler binding in E3.5 ICM, where SOX2 binds preaccessible enhancers, and its loss does not substantially affect chromatin opening. These preaccessible enhancers are in part opened by early-stage expressing TFs TFAP2C and NR5A2. Notably, settler binding of SOX2 can still exert impacts on gene expression, especially at sites with strong SOX2 motifs. Pioneer binding of SOX2 becomes widespread in E4.5 epiblast and naive ESCs (2i ESCs), where SOX2 is required for opening naive enhancers, and its binding sites enrich the OCT4-SOX2 motif. Finally, the pilot binding of SOX2 at many formative enhancers in 2i ESCs is insufficient for enhancer opening but poises enhancers for faster opening upon the conversion to formative pluripotency.

CONCLUSION: In this work, we dissected the SOX2-governed pluripotency regulatory network in mouse early embryos. These data revealed highly dynamic regulatory circuitry during pluripotency progression in vivo, especially when cells enter naive and formative pluripotency. We also found multifaceted pioneer factor–enhancer interactions that underpin the transition of pluripotency states. Finally, these results also identified a distinct "prepluripotency" state in E3.5 ICM, between totipotency and pluripotency. Prepluripotency features the potency to give rise to both epiblast and primitive endoderm, coexpression of multilineage TFs, and a primitive pluripotency network that lacks interdependence of master pluripotency TFs and their pioneer binding. How these master TFs acquire chromatin opening ability to establish a pluripotency network when cells enter naive pluripotency warrants future investigations. Hence, these data bridge the knowledge gap between in vivo development and in vitro cultured stem cells and pave the way for future studies to understand pluripotency and cell fate decision. ■



SOX2 regulatory circuitry during pluripotency progression reveals multifaceted master TF–enhancer interaction. SOX2 regulatory circuitry was dissected during the pluripotency transition in mouse early embryos. The prepluripotency state in E3.5 ICM is proposed to bridge totipotency to pluripotency. Three SOX2-enhancer interaction modes are shown: settler, pioneer, and pilot binding.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: xiewei121@tsinghua.edu.cn
†These authors contributed equally to this work.
Cite this article as L. Li *et al.*, *Science* **382**, eadi5516 (2023).
DOI: 10.1126/science.adi5516

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adi5516

RESEARCH ARTICLE

DEVELOPMENT

Multifaceted SOX2-chromatin interaction underpins pluripotency progression in early embryos

Lijia Li[1,2]†, Fangnong Lai[1,2]†, Xiaoyu Hu[1,2]†, Bofeng Liu[1,2]†, Xukun Lu[1,2], Zili Lin[3], Ling Liu[1,2], Yunlong Xiang[4], Tristan Frum[5,6], Michael A. Halbisen[5], Fengling Chen[1,2], Qiang Fan[1,2], Amy Ralston[5], Wei Xie[1,2]*

Pioneer transcription factors (TFs), such as OCT4 and SOX2, play crucial roles in pluripotency regulation. However, the master TF-governed pluripotency regulatory circuitry was largely inferred from cultured cells. In this work, we investigated SOX2 binding from embryonic day 3.5 (E3.5) to E7.5 in the mouse. In E3.5 inner cell mass (ICM), SOX2 regulates the ICM-trophectoderm program but is dispensable for opening global enhancers. Instead, SOX2 occupies preaccessible enhancers in part opened by early-stage expressing TFs TFAP2C and NR5A2. SOX2 then widely redistributes when cells adopt naive and formative pluripotency by opening enhancers or poising them for rapid future activation. Hence, multifaceted pioneer TF–enhancer interaction underpins pluripotency progression in embryos, including a distinctive state in E3.5 ICM that bridges totipotency and pluripotency.

The mammalian embryo undergoes several rounds of cell cleavage after fertilization to give rise to a blastocyst (*1*, *2*). In the mouse, blastocyst is specified into trophectoderm (TE), which eventually contributes to placenta, and inner cell mass (ICM) which subsequently differentiates into primitive endoderm (PrE) and epiblast (*1*–*3*). During this process, cells with pluripotency emerge and transit through several states that could be captured ex vivo, including the naive, formative, and primed pluripotency (*4*). Naive pluripotency is represented by the pluripotency state exhibited by naive embryonic stem cells (2i ESCs) that resemble preimplantation epiblast. After implantation, the epiblast cells acutely transit to the formative pluripotency state that manifests competence for both primordial germ cell (PGC) and somatic fate induction (*4*). This transition can be recapitulated by naive embryonic stem cells differentiating into epiblast-like cells (EpiLCs) (*5*). Finally, epiblast stem cells (EpiSCs), considered to recapitulate primed pluripotency, could be derived from E5.5 to E8 epiblast or ectoderm (*6*). Unlike formative pluripotent cells, EpiSCs have lost the competence to PGC induction (*5*).

Transcription factors (TFs) play central roles in development by means of precise spatiotemporal regulation of gene expression through action at cis-regulatory elements (*7*, *8*). Key master TFs, such as NANOG, SOX2, and OCT4, are essential to the pluripotency network (*9*). Among them, SOX2 is the earliest and only pluripotent TF known to be restricted to the inner cells in the mouse morula (16-cell), the ICM progenitors (*10*). *Sox2* deficiency in mice leads to epiblast formation failure and embryonic lethality shortly after implantation (*11*). Because of the limited research materials from embryos, the regulatory circuitry governed by master pluripotency factors has largely been inferred from cultured cell models (*9*, *12*, *13*), with limited TF regulomes being investigated in mouse blastocysts (e.g., NANOG) (*14*, *15*). Therefore, how these TFs regulate pluripotency and its transition in physiological conditions remains elusive. For example, *Sox2*-deficient embryos can still form blastocysts containing a morphologically normal ICM, which, however, cannot give rise to ESCs. Consistently, ESCs lose pluripotency when SOX2 is depleted (*11*, *16*). Furthermore, OCT4 and SOX2 are considered to be "pioneer factors" that can open inaccessible chromatin and recruit other TFs (*17*, *18*). How they potentially drive such pluripotency transition in vivo remains unknown. In this work, we applied CUT&RUN to capture the SOX2-chromatin binding from E3.5 ICM to E7.5 ectoderm, which covers the entire progression of pluripotency. Our results revealed that SOX2 engages enhancers through stage- and context-dependent action modes to regulate development programs. Furthermore, these data reveal a distinctive state and regulatory circuitry in E3.5 ICM that connects totipotency and pluripotency.

Genome-wide mapping of SOX2-chromatin binding in mouse early embryos during pluripotency progression

By identifying a SOX2 antibody to allow CUT&RUN analyses using as few as 200 cells (fig. S1, A to C), we examined SOX2-chromatin binding at stages when it is expressed, including E3.5 ICM from early and middle blastocyst, epiblast from E4.5 preimplantation embryos, E5.5 and E6.5 postimplantation embryos, and ectoderm dissected from E7.5 embryos (fig. S2, A and B; Fig. 1, A and B; and materials and methods), which represent states from the onset to the exit of pluripotency. For comparison, we also included 2i ESCs and EpiLCs, which resemble E4.5 epiblast and E5.5 epiblast, respectively (*19*).

We conducted several analyses to validate these datasets. (i) The CUT&RUN data were reproduced between replicates (Fig. 1C and fig. S2C). (ii) Globally, 82.8 to 92.0% of SOX2 peaks occupied intergenic and intragenic regions away from promoters (fig. S2D), consistent with TFs predominantly binding to enhancers (*8*). For convenience, we refer to SOX2-bound distal regions as putative enhancers or enhancers hereafter. An examination of known SOX2 target genes *Pou5f1* and *Nanog* (*12*, *16*, *20*) revealed SOX2 binding at their known enhancers (Fig. 1D). (iii) We generated *Sox2* maternal and zygotic knockout (mzKO) embryos and found that SOX2 binding was substantially diminished in mutant E3.5 ICM (Fig. 1B and fig. S2E). (iv) Across all stages, SOX2-bound putative enhancers were enriched for SOX2 motif and exhibited distinct motif enrichment compared with assay for transposase-accessible chromatin sequencing (ATAC-seq) (*21*) (Fig. 1E), which suggests that SOX2 binding is not simply dependent on chromatin accessibility. (v) Finally, by acute degradation of SOX2 in 2i ESCs (*22*) (fig. S3, A to C), we observed that SOX2-targeted enhancers near ICM-specific genes *Upp1* and *Spic* exhibited SOX2-dependent transcriptional regulatory activity (fig. S3D). Taken together, these analyses suggest that we have identified bona fide SOX2 binding targets in early embryos.

SOX2 binding in E3.5 ICM manifests a state that is distinct from previously defined pluripotent states and stem cells

Global analysis of SOX2 distal binding sites showed that SOX2 binding in E3.5 ICM and E4.5 epiblast was highly divergent from that at other developmental stages (Fig. 1, C, F, and G), as confirmed by correlation analyses (fig. S4A). SOX2 binding in E3.5 ICM was also distinct from that in 2i ESCs and EpiLCs. An analysis of gene expression and chromatin accessibility

[1]Center for Stem Cell Biology and Regenerative Medicine, MOE Key Laboratory of Bioinformatics, School of Life Sciences, New Cornerstone Science Laboratory, Tsinghua University, Beijing 100084, China. [2]Tsinghua-Peking Center for Life Sciences, Beijing 100084, China. [3]College of Animal Science and Technology College, Beijing University of Agriculture, Beijing 102206, China. [4]Department of Cell Biology and Genetics, School of Basic Medical Sciences, Chongqing Medical University, Chongqing 400016, China. [5]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA. [6]Department of Internal Medicine, Gastroenterology, University of Michigan Medical School, Ann Arbor, MI 48109, USA.
*Corresponding author. Email: xiewei121@tsinghua.edu.cn
†These authors contributed equally to this work.

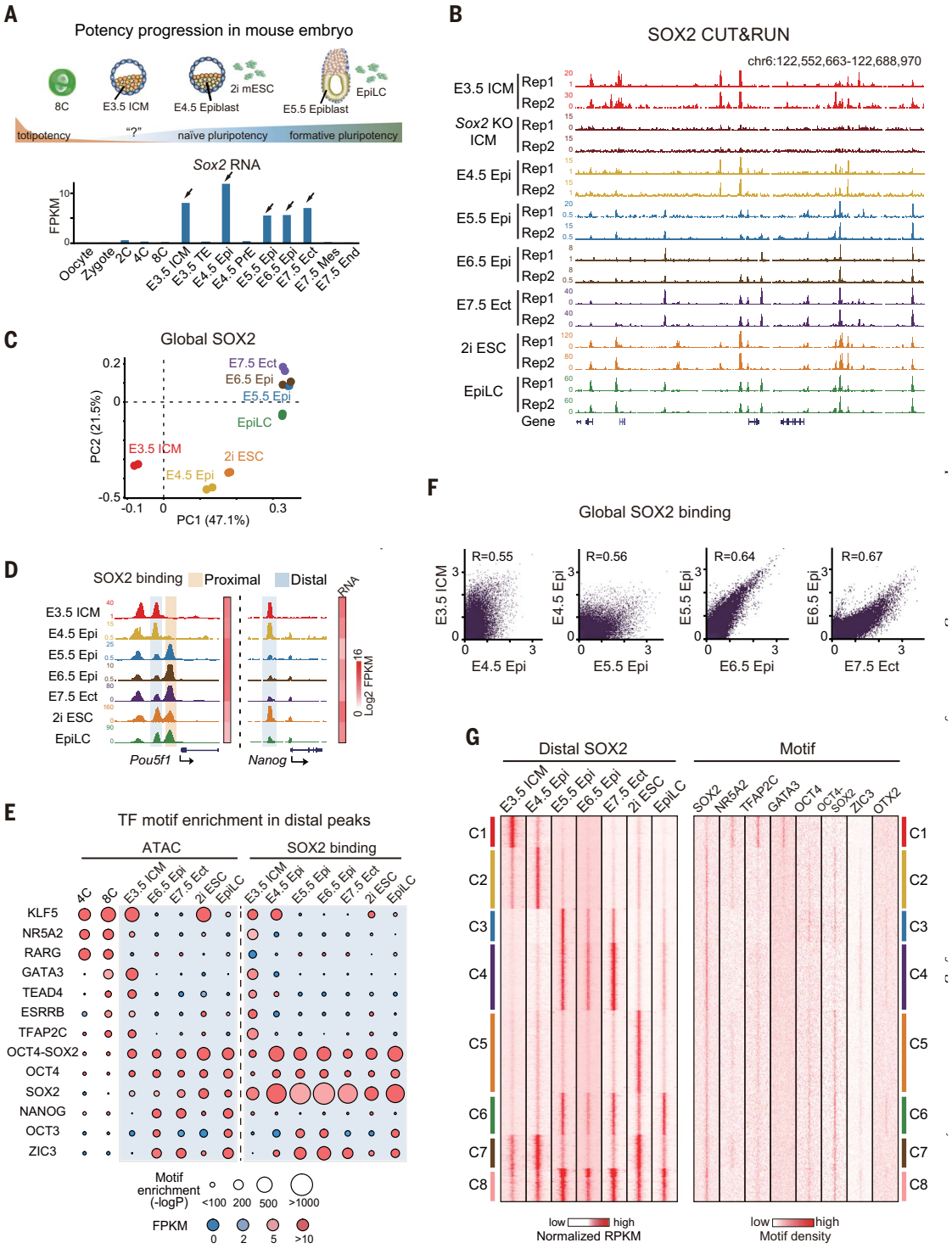


Fig. 1. The dynamics of SOX2-chromatin binding during pluripotency progression in vivo. (**A**) Developmental stages and related *Sox2* expression are shown. Arrows indicate lineages where SOX2 binding was profiled. FPKM, fragments per kilobase of transcript per million mapped fragments. (**B**) UCSC browser view showing SOX2 CUT&RUN signals in E3.5 ICM; *Sox2* mzKO ICM; E4.5, E5.5, and E6.5 epiblast (Epi); E7.5 ectoderm (Ect); 2i ESC; and EpiLC (two biological replicates). (**C**) Principal components analysis (PCA) of global SOX2 binding enrichment in mouse early lineages and stem cells. (**D**) UCSC browser views and heatmaps showing SOX2 enrichment and gene expression of representative genes, respectively. Proximal and distal binding sites are shaded. (**E**) TF motifs identified from distal ATAC-seq peaks and SOX2 binding peaks. Sizes of circles indicate levels of −log *P* value. Expression levels of TFs are color coded. (**F**) Scatter plots showing global SOX2 binding correlation between consecutive stages. (**G**) Heatmaps showing the SOX2 binding signals and TF motif densities at stage-specific and shared distal SOX2 binding peaks. C1 to C8, cluster 1 to cluster 8; RPKM, reads per kilobase per million mapped reads.

(*21*, *23*) independently supported the distinctive state of E3.5 ICM (fig. S4B). E3.5 SOX2 binding sites preferentially harbored motifs of NR5A2, GATA, and TFAP2C, which were also enriched in accessible chromatin at the earlier stage [8-cell (8C)] but not the OCT4 and the composite OCT4-SOX2 motifs (Fig. 1E). These data suggest a poised state of E3.5 ICM that coexpresses not only master pluripotency genes (*Sox2*, *Oct4*, and *Nanog*) but also PrE genes (*Gata6*) and early-stage TFs (*Nr5a2* and *Tfap2c*) (*3*, *24*, *25*) (fig. S4C).

SOX2 then underwent substantial redistribution when cells from E3.5 ICM entered naive pluripotency (E4.5 epiblast and 2i ESC) and again when cells entered formative pluripotency (E5.5 epiblast and EpiLC) (Fig. 1, C, F, and G). By contrast, stages from E5.5 epiblast onward (E6.5 epiblast and E7.5 ectoderm) were clustered closely, which suggests a more gradual formative-to-primed pluripotency transition. Because the pluripotency transition is a continuum, E5.5 epiblast may represent the formative state, whereas E6.5 epiblast may be transitionary between formative and primed pluripotency (*4*). Given their similarity, we mainly focused on E5.5 epiblast in subsequent

analyses. SOX2 binding showed a moderate transition from E6.5 epiblast to E7.5 ectoderm. A small number of binding sites were newly established in E7.5 ectoderm, accompanied by increased ATAC-seq and H3K27ac signals (fig. S4D). Only a small subset (8.8%) of SOX2 distal peaks were shared among all seven cell types (Fig. 1G, "C8"). Genes related to blastocyst and ICM formation were enriched near SOX2 binding sites in E4.5 epiblast (fig. S4E, "C2"). Neural tube patterning– and stem cell maintenance–associated genes were enriched near enhancers shared in E5.5 epiblast, ectoderm, and EpiLCs (fig. S4E, "C6"). These data demonstrate highly dynamic SOX2 binding during pluripotency progression.

SOX2 is required for the ICM-TE lineage program in E3.5 ICM

To investigate whether SOX2 binding in E3.5 ICM is linked to gene expression, we compared SOX2 binding from E3.5 ICM with that from E4.5 epiblast. E3.5 ICM–specific SOX2 binding preferentially resided near E3.5 ICM–specific genes (e.g., *Gata6*), whereas E4.5 epiblast–specific SOX2 binding tended to occur near E4.5 epiblast–activated genes (e.g., *Lefl*) (Fig. 2A, "RNA"). We then performed single-cell RNA sequencing (scRNA-seq) in wild-type (WT) and *Sox2* mzKO E3.5 ICM (Fig. 2, B and C). The coexpression of the epiblast gene *Nanog* and the PrE gene *Gata6* in most cells (fig. S5A) confirmed that the epiblast and PrE segregation has not started in E3.5 ICM (*3*). The mutant cells were globally separated from WT cells (Fig. 2B), with 510 genes up-regulated and 623 genes down-regulated in the mutant cells (fig. S5B). In total, 239 of the up-regulated genes were TE-enriched genes (such as *Gata3*, *Krt8*, *Eomes*, and *Id2*), whereas 99 of the down-regulated genes were ICM specific (Fig. 2D and fig. S5C). The latter included not only epiblast-enriched genes, such as *Spic*, *Fgf4*, *Utf1*, and *Upp1*, but also PrE genes, such as *Sox17* and *Pdgfra* (Fig. 2E). Notably, the expression of *Pou5f1* and *Nanog* was not significantly affected as previously reported (*10*) (Fig. 2C and fig. S5A). Therefore, although the blastocyst morphology appears unaffected (*10*, *11*), the ICM-TE transcription program is severely impaired in *Sox2* mutant ICM.

The OCT4 and SOX2-OCT4 composite motif enrichment within SOX2 binding sites was low in E3.5 ICM (Fig. 1E). They became enriched in E4.5 epiblast (Fig. 1G), which indicates that the cooperativity of these two pluripotency factors may only become mature when cells enter naive pluripotency. Although altered ICM-TE expression was also reported in *Oct4* KO E3.5 ICM (*26*) (fig. S6A), the gene expression changes between *Oct4* KO and *Sox2* KO E3.5 ICM overall showed a weak correlation (correlation coefficient $r = 0.15$) (fig. S6B). Only 96 down-regulated genes (e.g., ICM-specific genes *Fgf4*, *Spp1*, and *Spic*) and 95 up-regulated genes (e.g., TE-specific genes *Dppa1*, *Gata3*, and *Krt18*) were shared in *Sox2* and *Oct4* KO ICM (fig. S6C), which indicates that OCT4 and SOX2 have both shared and specific functions at this stage.

SOX2 binds preaccessible chromatin and is dispensable for enhancer opening in E3.5 ICM

We next investigated how enhancers were affected in *Sox2* mzKO E3.5 ICM. However, ATAC-seq showed few global changes at the SOX2-bound sites (Fig. 3, A and B). A close examination revealed that 75.5% of SOX2-occupied sites in ICM—either E3.5 ICM–specific or E3.5 to E4.5 shared—were already accessible ("preaccessible," or "preaccess") in 8C embryos (Fig. 3, A and C, "*Fgf4*") when *Sox2* expression was still undetectable (Fig. 1A and fig. S2A). SOX2 binding was diminished at preaccessible sites in *Sox2* KO ICM, which confirms that these are bona fide signals (fig. S6D). Although the remaining SOX2-occupied sites became accessible after the 8C stage (opened de novo after *Sox2* expression, or "de novo") (Fig. 3, A and C), SOX2 was largely dispensable for accessibility of both preaccessible and de novo enhancers (Fig. 3, A and B). In fact, one-third of the de novo sites already showed deoxyribonuclease (DNase) I hypersensitive sites sequencing (DNase-seq) signals (*27*) in morula at the 16C stage (fig. S6E). By contrast, E4.5 epiblast–specific SOX2-bound enhancers required SOX2 for opening (Fig. 3C, "*Pdgfc*," and Fig. 3D), which displayed stronger SOX2 motif enrichment compared with E3.5 ICM–specific enhancers (Fig. 3A). Hence, SOX2 preferentially resided at preaccessible chromatin sites in E3.5 ICM. We referred to this TF binding mode, in which a TF binds enhancers that are preaccessible and its loss does not substantially affect chromatin opening, as "settler binding" (*28*). SOX2 then adopted an indispensable role in opening E4.5 epiblast–specific enhancers, which we referred to as "pioneer binding." We note that settler and pioneer here refer to different TF binding modes, which may come from the same TF (e.g., SOX2), rather than different classes of TFs.

We investigated whether such settler binding of SOX2 is associated with gene expression. We found that genes down-regulated (but not the up-regulated genes) in *Sox2* KO E3.5 ICM preferentially resided near enhancers that were bound by SOX2 in both E3.5 ICM and E4.5 epiblast (Fig. 3B, "RNA," "shared," and "preaccess"). These enhancers harbored strong SOX2 motif enrichment similar to E4.5 epiblast–specific enhancers (Fig. 3A, "SOX2 motif"). By contrast, SOX2 settler binding at E3.5 ICM–specific enhancers, which exhibited much weaker SOX2 motif enrichment, appeared to have minimal impacts on nearby genes (Fig. 3B, "E3.5 ICM specific" and "preaccess"). Therefore, settler binding of SOX2 at enhancers with strong SOX2 motifs is associated with gene activation.

Early-stage expressing TFs are responsible for opening preaccessible SOX2 binding sites

We next investigated which TFs might be responsible for opening preaccessible SOX2 binding sites. Both E3.5 ICM–specific and E3.5 ICM–E4.5 epiblast shared SOX2-bound enhancers enriched for motifs of TFAP2C, NR5A2, and GATA (Fig. 4A). *Tfap2c* and *Nr5a2* show highest expression at the 4C to 8C stages (fig. S4C), and both are essential for embryogenesis (*29*–*31*). TFAP2C regulates cell polarity starting from the 2C stage before being restricted to extraembryonic lineages in the late blastocyst (*32*, *33*). NR5A2, a pioneer factor (*34*), is essential for naive pluripotency (together with ESRRB) in 2i ESCs (*35*) and can regulate the ICM program as early as the 8C stage (*21*). KO of *Nr5a2* arrested embryos at the morula stage (*36*, *37*).

We then examined the occupancy of TFAP2C and NR5A2 (*37*) in 8C embryos using CUT&RUN. TFAP2C and NR5A2 extensively resided at preaccessible sites but not at E4.5 epiblast–specific SOX2 binding sites (Fig. 4, A and B). Among 5578 distal SOX2-bound peaks in E3.5 ICM, 1571 and 1281 sites were already occupied by TFAP2C and NR5A2, respectively, at the 8C stage before *Sox2* expression (Fig. 4C). To determine whether these TFs can open these enhancers, we conducted ATAC-seq in *Tfap2c* mzKO 8C embryos. Depletion of TFAP2C preferentially reduced accessibility of preaccessible enhancers specifically bound by TFAP2C (Fig. 4D and fig. S7A). Overall, TFAP2C contributed to opening 35% preaccessible enhancers (SOX2-bound in E3.5 ICM) (Fig. 4E). A similar but less widespread effect was observed for NR5A2 at 8C, as *Nr5a2* knockdown decreased chromatin accessibility of 23% of preaccessible enhancers (Fig. 4E and fig. S7B). Together, TFAP2C and NR5A2 accounted for opening 49% preaccessible enhancers (with 9% regulated by both) (Fig. 4E). The rest of the preaccessible sites may be opened by other early-stage TFs (e.g., GATA) (Fig. 4A).

In 8C embryos, NR5A2 was preferentially required for opening chromatin at the 8C-specific binding sites but not regions bound by NR5A2 starting from the 2C stage (fig. S7C, "8C ATAC"), which suggests that these 2C-specific sites were also preaccessible for NR5A2. Indeed, *Nr5a2* knockdown did not substantially affect global chromatin accessibility in 2C embryos (fig. S7C, "2C ATAC"), resembling SOX2's role in opening enhancers in E4.5 epiblast but not E3.5 ICM. Therefore, these data indicate that both SOX2 and NR5A2 tend to initially exhibit settler binding at preaccessible enhancers that are likely opened

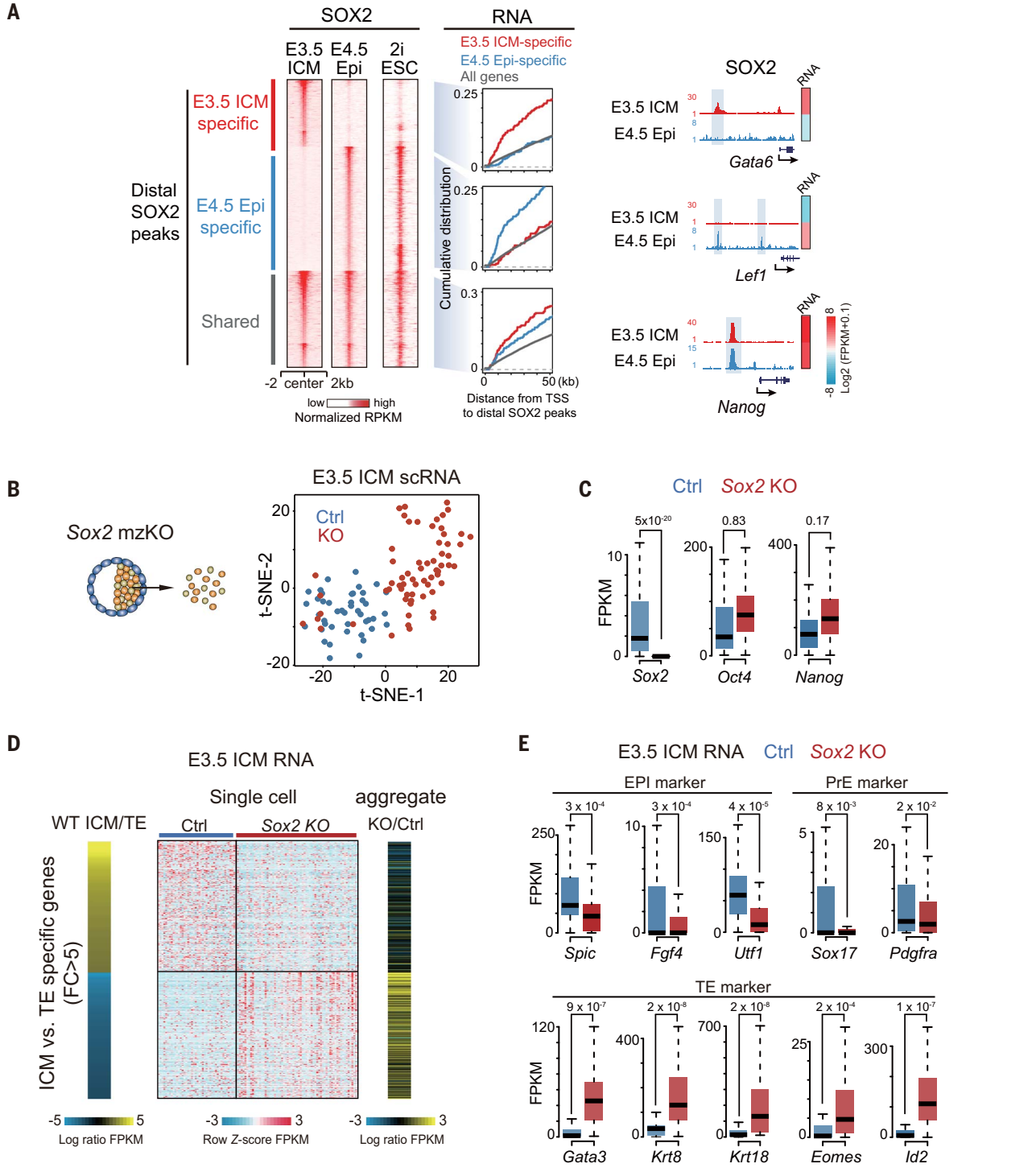


Fig. 2. SOX2 regulates ICM-TE lineage gene expression in E3.5 ICM. (**A**) (Left) Heatmaps showing enrichment of SOX2 binding signal at E3.5 ICM–specific (n = 1534), E4.5 epiblast–specific (n = 2848), and shared (n = 2223) SOX2 binding peaks. (Middle) The cumulative distribution of E3.5 ICM–specific, E4.5 epiblast–specific, and all genes shows the fraction of genes within defined distances (x axis) between their transcription start sites (TSSs) to the nearest distal SOX2 binding peaks. (Right) UCSC browser views show SOX2 binding enrichment of representative genes. Gene expression is shown in heatmaps. (**B**) A t-distributed stochastic neighbor embedding (t-SNE) plot for scRNA-seq of control (blue) and *Sox2* mzKO (red) ICM at E3.5. (**C**) Box plots showing *Sox2*, *Oct4*, and *Nanog* gene expression in control and *Sox2* mzKO ICM single cells, with P values (t test, one-sided) indicated. (**D**) (Left) Heatmaps showing the fold changes between ICM and TE gene expression. (Middle) The expression in control (Ctrl) and *Sox2* mzKO single cells (row z-score normalized) is mapped. (Right) The aggregated KO/control gene expression ratios (across all single cells) are shown. (**E**) Box plots showing expression of representative epiblast (EPI), PrE, and TE markers in control and *Sox2* mzKO ICM single cells, with P values (t test, one-sided) indicated.

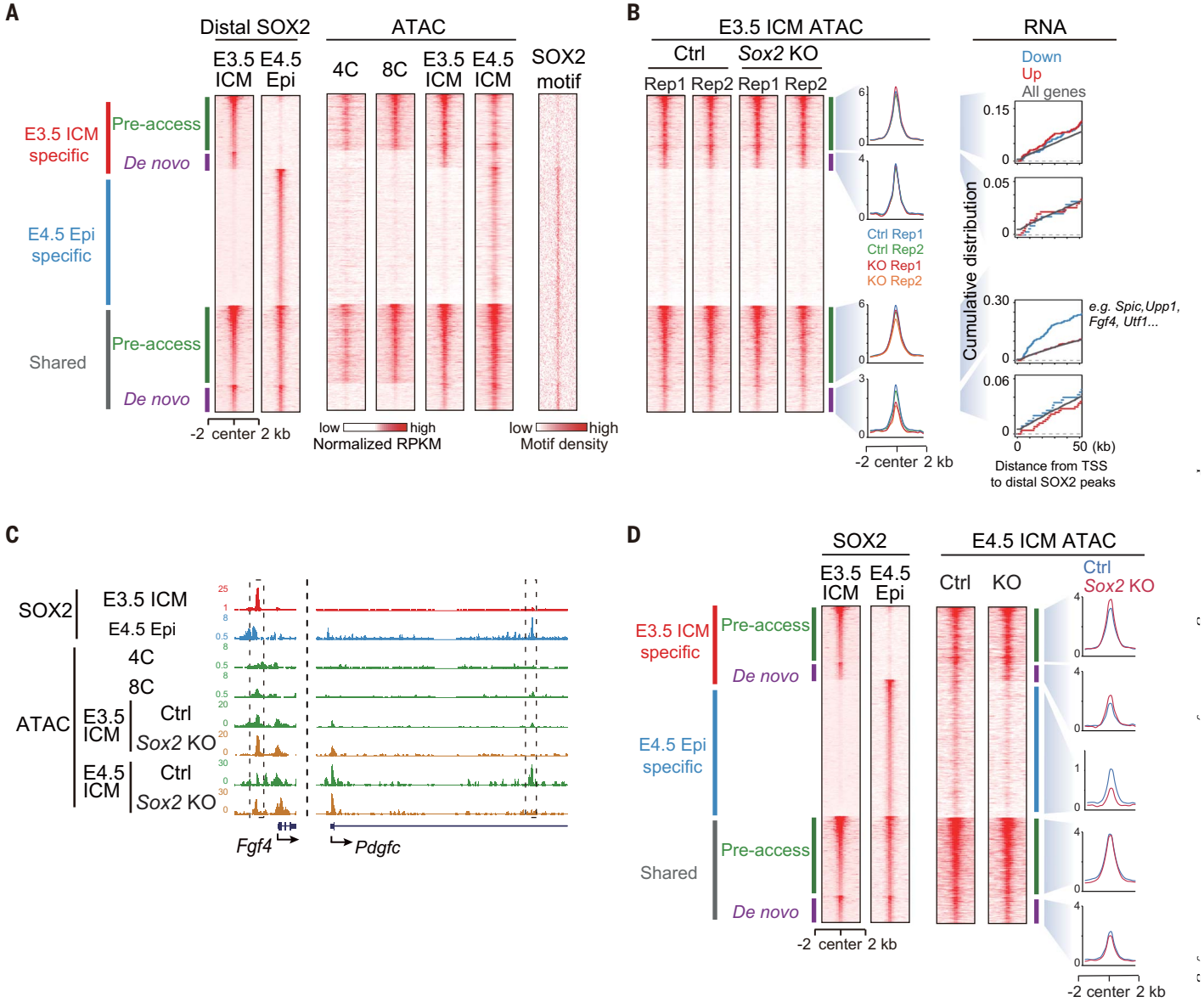


Fig. 3. SOX2 is dispensable for global enhancer accessibility in E3.5 ICM. (**A**) (Left) Heatmaps showing enrichment of SOX2 binding signal, ATAC-seq signals, and SOX2 motif density (number of motifs per base pair) at SOX2 binding peaks. The E3.5 ICM–specific peaks are further clustered into preaccessible (n = 1177) and de novo (n = 357) peaks on the basis of their accessibility states in 8C embryos. The shared peaks are also clustered into preaccessible (n = 1658) and de novo (n = 565) peaks. (**B**) (Left) Heatmaps and average plots showing enrichment of ATAC-seq signals in control and *Sox2* mzKO E3.5 ICM at SOX2 binding peaks. (Right) The cumulative distributions of down-regulated, up-regulated, and all genes in *Sox2* mzKO E3.5 ICM show the fraction of genes within defined distances (x axis) from their TSSs to the nearest SOX2 binding peaks. (**C**) UCSC browser views showing SOX2 binding (dashed boxes) and ATAC-seq enrichment of representative genes. (**D**) Heatmaps and average plots showing enrichment of SOX2 binding and ATAC-seq signals in control and *Sox2* mzKO E4.5 ICM at SOX2 binding peaks.

by other TFs before adopting a more active role in opening enhancers at later stages.

SOX2 mediates global enhancer opening in naive pluripotent cells

Although SOX2 was required to open E4.5 epiblast–specific enhancers, the failure to open these enhancers in *Sox2* mzKO mutants could also be indirectly caused by earlier defective development (Fig. 2D). Therefore, we investigated the dependence of SOX2 for opening enhancers in 2i ESCs, which resemble E4.5 epiblast (*4*, *38*). We confirmed that E4.5 epiblast–specific, but not E3.5 ICM–specific, enhancers were preferentially bound by SOX2 in 2i ESCs (Fig. 2A). About 67.3% of SOX2 binding sites in E4.5 epiblast were recaptured in 2i ESCs. We then focused on SOX2 binding sites that were shared between 2i ESCs and E4.5 epiblast (Fig. 5A). RNA-seq showed that transcriptome defects were minimal at 12 hours upon SOX2 degradation but became apparent after 24 hours (fig. S8A). Differentially expressed genes (DEGs) within 48 hours (521 down-regulated and 462 up-regulated) upon SOX2 degradation were identified (fig. S8A). Compared with DEGs in *Sox2* KO E3.5 ICM, only 52 down-regulated genes were shared, including 12 epiblast-specific genes (e.g., *Klf2*, *Fgf4*, and *Etv5*). E3.5 ICM–specific down-regulated genes contained more PrE genes (e.g., *Gata4*, *Pdgfra*, and *Sox17*), whereas ESC-specific down-regulated genes included more epiblast markers (e.g., *Esrrb*, *Klf4*, and *Nanog*)

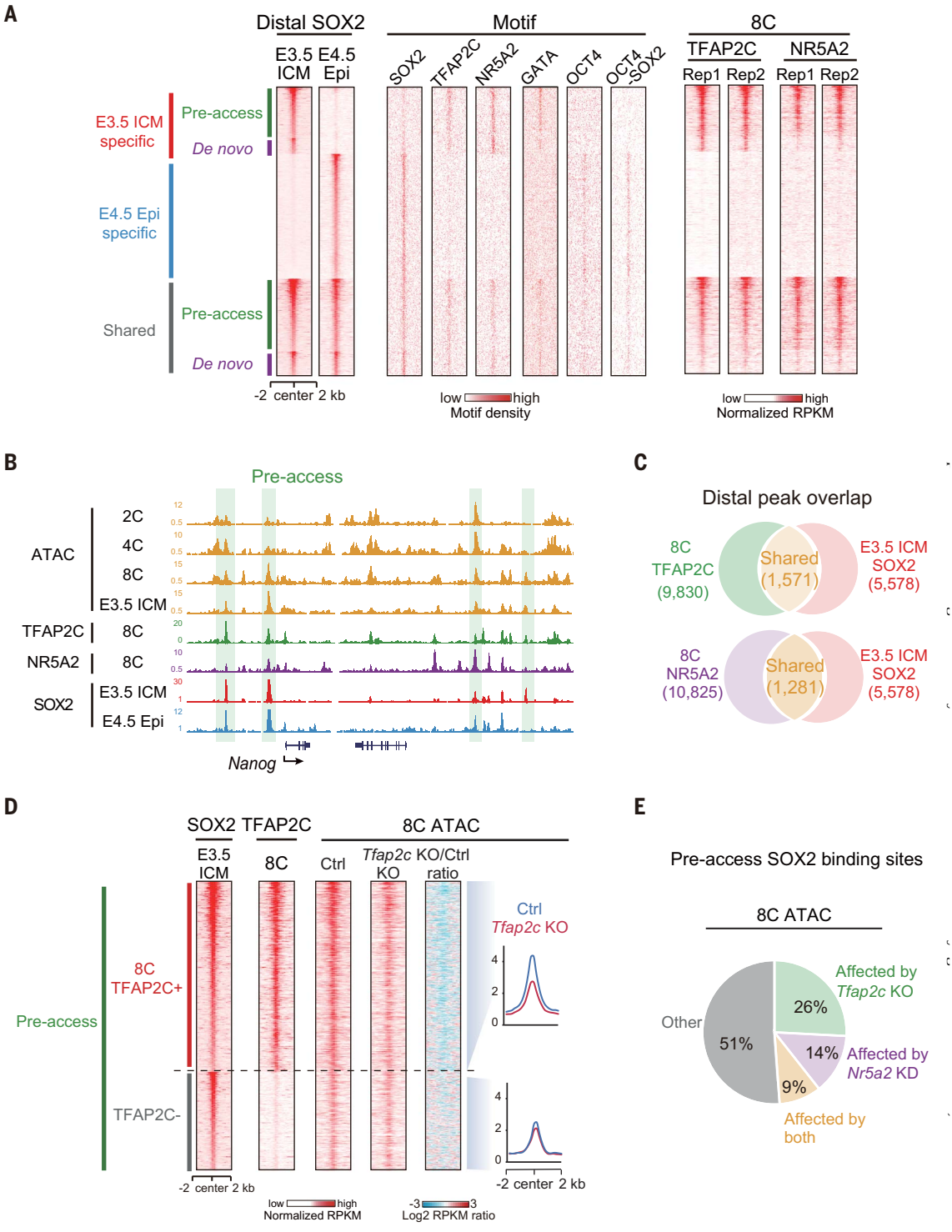


Fig. 4. Early-stage expressing TFs are responsible for opening the preaccessible SOX2 binding sites. (**A**) Heatmaps showing enrichment of SOX2 binding signals in E3.5 ICM and E4.5 epiblast, motif densities, and TFAP2C and NR5A2 binding signals in 8C embryos at SOX2 binding peaks. (**B**) UCSC browser views showing ATAC-seq enrichment and TFAP2C, NR5A2, and SOX2 binding signals of representative regions. Preaccessible binding sites are shaded. (**C**) Venn diagrams showing the overlap between distal TFAP2C (top) or NR5A2 (bottom) binding peaks in 8C embryos and SOX2 binding peaks in E3.5 ICM. (**D**) Heatmaps showing SOX2 binding in E3.5 ICM, TFAP2C binding signals in 8C embryos, ATAC-seq enrichment in control and *Tfap2c* mzKO 8C embryos, and the ratios between *Tfap2c* mzKO and control 8C embryos at the preaccessible E3.5 ICM SOX2 binding peaks. Average plots of ATAC-seq signals are shown. (**E**) Pie chart showing the percentages of *Tfap2c* KO affected, *Nr5a2* knockdown (KD) affected, and both affected ATAC-seq signal at the preaccessible SOX2 binding peaks.

(Fig. 5B and fig. S8B). Hence, SOX2 is required for broad expression of pluripotency marker genes in 2i ESCs but not in E3.5 ICM. On the other hand, several key TE markers were up-regulated in both cell types (e.g., *Gata2*, *Id2*, and *Krt8/18*) (>Fig. 5B). These results support a distinct state of E3.5 ICM from naive ESCs, in part reflected by its primitive pluripotency network and broader potency toward both epiblast and PrE.

We then performed ATAC-seq and H3K27ac chromatin immunoprecipitation followed by sequencing (ChIP-seq) at day 0.5 (D0.5, or 12 hours) after SOX2 degradation, when transcription perturbation was still small (fig. S8A), to minimize secondary effects. About 59.8% of SOX2-bound ATAC peaks were already lost compared with only 36.8% of SOX2-unbound enhancers (Fig. 5C and fig. S9A; see Fig. 5D for example). SOX2-dependent enhancers showed relatively stronger SOX2 binding and SOX2 motif enrichment (Fig. 5, E and F). Enhancers with both SOX2 and OCT4 motifs showed the highest SOX2 binding and SOX2 dependency (fig. S9B). Moreover, the putative target genes of the SOX2-dependent enhancers (materials and methods) were down-regulated at an earlier time point (D0.5) upon

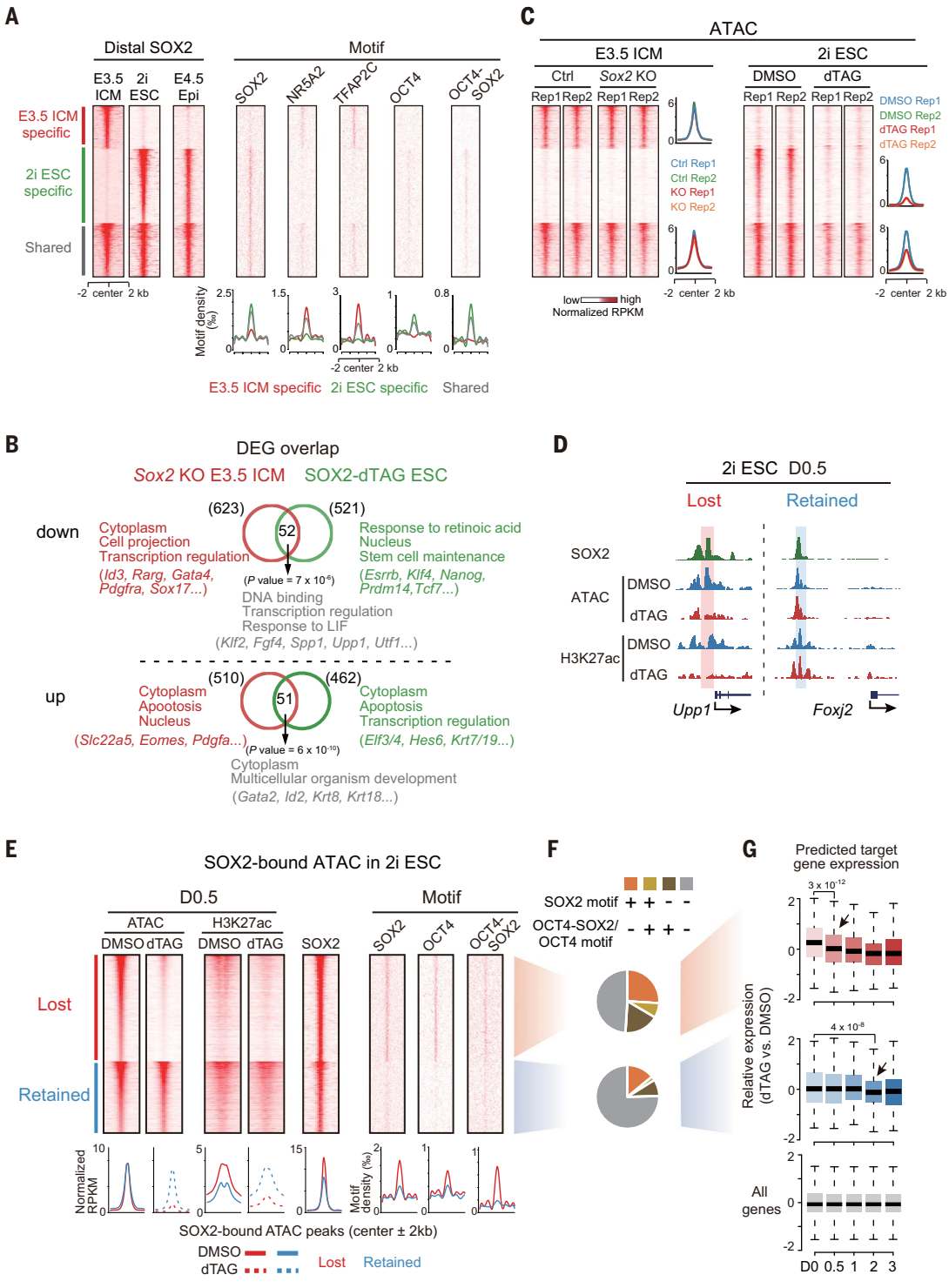


Fig. 5. Distinct roles of SOX2 in gene expression and enhancer regulation in E3.5 ICM and 2i ESCs. (**A**) Heatmaps showing enrichment of SOX2 binding signals and TF motif densities at E3.5 ICM–specific, 2i ESC–specific (also E4.5 epiblast FPKM > 1), and shared (also E4.5 epiblast FPKM > 1) distal SOX2 binding peaks, with the average plots of TF motif densities shown below. (**B**) Venn diagrams showing the overlap of down- and up-regulated genes between *Sox2* KO (versus control) E3.5 ICM and dTAG [versus dimethyl sulfoxide (DMSO)–treated] 2i ESCs. Gene ontology results and example genes are shown. *P* values (hypergeometric distribution) are shown. (**C**) Heatmaps showing enrichment of ATAC-seq signals in control and *Sox2* mzKO E3.5 ICM and in DMSO- and dTAG-treated (for 12 hours) 2i ESCs. The average plots of ATAC-seq signal are shown. (**D**) UCSC browser views showing enrichment of SOX2 binding, ATAC-seq, and H3K27ac signals in DMSO- and dTAG-treated 2i ESCs of representative genes. The lost and retained ATAC-seq peaks are shaded. (**E**) Heatmaps showing enrichment of ATAC-seq, H3K27ac, SOX2 binding signals, and TF motif densities at lost and retained ATAC-seq peaks in DMSO- or dTAG-treated 2i ESCs. The average enrichment is shown below. (**F**) Pie charts showing the percentages of peaks with SOX2 motifs only, both SOX2 and OCT4-SOX2/OCT4 motifs, OCT4-SOX2/OCT4 motifs, and neither motif at the peaks. (**G**) The relative expression of predicted target genes (row *z*-score normalized) of the ATAC-seq peaks and all genes (control), with *P* values (*t* test, two-sided) indicated on top. The arrows indicate the time when a significant decrease of gene expression is detected.

SOX2 loss compared with targets of SOX2-independent enhancers, which were down-regulated around day 2 (Fig. 5G), supporting more direct SOX2 impacts. SOX2-independent enhancers showed comparable ATAC-seq and even stronger H3K27ac signals compared with SOX2-dependent enhancers in WT cells (Fig. 5E), which suggests that they may be opened by other TFs, although the motif analysis did not reveal obvious candidates (fig. S9C). In sum, these data suggest that in the naive pluripotent state, SOX2 opens enhancers—preferentially those with SOX2 or OCT4 motifs—in a pioneer binding mode. Together with the dispensability of SOX2 for opening global enhancers in E3.5 ICM, this result is in line with the finding that *Sox2*-null embryos can give rise to ICM but not to ESCs (*11*).

The second acute global binding transition underlies the essential role of SOX2 for formative pluripotency induction

Sox2-null embryos die shortly after implantation (*11*). Coincidently, SOX2 binding sites

showed a second major relocalization from E4.5 epiblast to E5.5 epiblast (fig. S10A), which suggests an involvement of SOX2 during the naive-to-formative pluripotency transition. In this process, whereas pluripotency genes *Sox2* and *Oct4* continue to be expressed, naive pluripotency genes (e.g., *Nanog*, *Tbx3*, and *Tbx20*) are repressed, and postimplantation epiblast genes (e.g., *Sall2*, *Fgf5*, and *Fgf15*) are activated (*5*). Accordingly, the motifs of SOX2 and OCT4 were enriched at both E4.5 epiblast– and E5.5 epiblast–specific SOX2-bound enhancers. By contrast, E5.5 epiblast–specific sites were enriched for motifs of ZIC3 and OTX2 (fig. S10A)—two TFs up-regulated during formative pluripotency induction, which mediate the naive-to-formative transition (*39*–*42*).

Because early lethality precludes the study of SOX2's role during the naive-to-formative pluripotency transition in vivo, we used the 2i ESC–to–EpiLC conversion ex vivo model (*5*, *43*). SOX2 binding in EpiLCs recapitulated that in E5.5 epiblast, with its binding sites enriched for the motifs of OTX2 and ZIC3 (fig. S10A). In fact, the co-occupancy of SOX2 with OCT4, ZIC3, and OTX2 was observed in EpiLCs (fig. S10B). After SOX2 depletion, global gene expression transition from naive-to-formative pluripotency was severely impaired (Fig. 6, A and B). About 57.3% (425 of 742) naive genes failed to be properly repressed (fig. S11A, left, "Down dependent"), and 78.2% (453 of 579) formative genes showed defective activation, including marker genes *Pou3f1*, *Fgf15*, *Dnmt3a/b*, *Zic3*, and *Otx2* (fig. S11A, left, "Up dependent," and fig. S11, B and C). The differentiation defects are likely a result of a systematic failure of the transcription program because they cannot be rescued by reintroducing *Otx2* or *Zic3* alone (fig. S11D). Supporting a direct role in gene activation, SOX2 in EpiLCs preferentially occupied enhancers near SOX2-dependent formative genes (fig. S11A, right). To probe how enhancers were globally affected upon the loss of SOX2, we performed ATAC-seq and H3K27ac ChIP-seq during the ESC-to-EpiLC transition. The decommissioning of ESC-specific enhancers (indicated by the loss of ATAC-seq peaks) upon differentiation was largely unaffected by SOX2 depletion (fig. S12A). However, 85% of SOX2-bound, newly established EpiLC-specific enhancers (compared with 61% of SOX2-unbound enhancers) failed to be properly established (Fig. 6C). To further validate the direct function of SOX2 at these enhancers, we tested SOX2-bound enhancers near *Otx2* and *Zic3*—two SOX2-dependent formative genes in EpiLCs (fig. S12B). All five enhancers that we tested drove strong reporter activities, with four being SOX2 dependent (fig. S12B). In sum, SOX2 is essential for the naive-to-formative pluripotency conversion and is required for the activation of formative enhancers.

Prebinding of SOX2 is insufficient to open enhancers but correlates with faster future enhancer opening

We hypothesized that SOX2-dependent formative-specific enhancers were activated by SOX2 through pioneer binding. However, we found that 54% of these enhancers were already prebound by SOX2 in 2i ESCs, despite still being inaccessible when assayed by ATAC-seq (Fig. 6D, "*Dnmt3b*," and Fig. 6E, "prebinding"). SOX2 binding at these sites further increased upon differentiation to EpiLCs. Prebinding of SOX2 at these enhancers was also observed in E4.5 epiblast in vivo (fig. S13A, "SOX2 in vivo"). The remaining 46% acquired SOX2 binding during differentiation, consistent with pioneer binding (Fig. 6D, "*Fgf15*," and Fig. 6E, "pioneer binding"). We investigated whether these enhancers with SOX2 prebinding and pioneer binding exhibit functional differences. Predicted target genes of both groups were activated with similar kinetics upon 2i ESC–to–EpiLC differentiation, in a SOX2-dependent manner (Fig. 6D, "Lost"). Thus, both SOX2 prebinding and pioneer binding were required for gene activation. By contrast, SOX2 was not essential to activate targets of SOX2-independent formative enhancers (Fig. 6D, "Retained").

We then sought to identify features that distinguish prebinding and pioneer binding enhancers. Both classes were depleted of H3K4me1 in ESCs, which indicates that they were not the classic "poised" enhancers (*44*) (fig. S13A). However, pioneer binding sites were enriched for both the SOX2 motif and the OTX2 motif, suggesting cooperative binding upon formative pluripotency induction (Fig. 6, F and G). Prebinding sites were more enriched for the SOX2 motif, which raises the possibility that such a strong motif is sufficient to recruit SOX2 in 2i ESCs, which may lower the threshold of enhancer activation to compensate for the weak OTX2 motif. We found that prebinding enhancers became accessible faster than pioneer binding enhancers during formative induction (Fig. 6E and fig. S13B), an observation also reproduced when analyzing H3K27ac (fig. S13C). Finally, OTX2 binding showed comparable increases in the two groups (fig. S13A, "OTX2"). We speculate that the increased binding of OTX2 in the prebinding enhancer group lacking its motif could be facilitated by other factors, such as SOX2. These data indicate that although prebinding of SOX2 is insufficient to open formative enhancers, it may poise those with weak formative TF motifs for faster future opening.

SOX2 prebinds germ layer enhancers in E5.5 epiblast

The fact that SOX2 can prebind and possibly poise enhancers for future activation prompted us to explore whether this observation can be extended to other developmental processes. During gastrulation, SOX2 is required to drive neural ectoderm differentiation (*45*). Many enhancers in ectoderm are already primed in mouse epiblast (*23*, *46*), consistent with the model that ectoderm is a default differentiation lineage from epiblast (*47*). By identifying the putative enhancers specific to epiblast and three germ layers using ATAC-seq data (*23*), we found that SOX2 occupied not only epiblast-specific but also 45% of ectoderm-specific enhancers in E5.5 epiblast (fig. S14, A and B). Moreover, SOX2 preferentially resided near both epiblast-specific and ectoderm-specific genes in E5.5 epiblast (fig. S14C). Upon the transition to ectoderm, SOX2 binding at ectoderm-specific enhancers was strengthened, whereas SOX2 binding at epiblast-specific enhancers was lost (fig. S14A). Hence, SOX2 prebinds a subset of developmentally regulated enhancers, which supports the notion that SOX2 functions as a lineage specifier toward ectoderm during gastrulation, and that formative pluripotence installs competence for somatic lineage specification (*4*). The prebinding of SOX2 resembles the reported binding of pioneer TFs and nonpioneer "bookmarking" TFs to regulatory elements before gene activation (*48*), but this binding does not immediately create open chromatin before receiving further differentiation cues. Therefore, we referred to such prebinding as "pilot binding," which further demonstrates the flexibility and versatility of pioneer factors in different cellular contexts. In fact, a quantitative analysis of enhancers for their SOX2 binding and chromatin accessibility across developmental stages revealed that different modes of SOX2-chromatin interactions are under constant transition (Fig. 7A). These distinct binding actions likely depend on cell type–specific cooperative TFs as well as genetic and epigenetic contexts.

Discussion

Tremendous progress has been achieved to understand the molecular circuitry underlying pluripotency regulation using stem cell models. How master TFs guide pluripotency progression in vivo remains poorly understood. In this work, by profiling the chromatin binding of SOX2 in mouse early embryos, we found a chromatin state and transcription circuitry for E3.5 ICM that differs from all other pluripotent states. The potency of E3.5 ICM exceeds pluripotency because it can give rise to epiblast and PrE. Moreover, E3.5 ICM shows distinct transcriptome and chromatin accessibility (fig. S4B), which likely reflect the coexpression of pluripotency factors, early-stage TFs, and extraembryonic lineage TFs. The pluripotency master TF–mediated regulatory network appears still at a primitive stage in E3.5 ICM, as supported by multiple

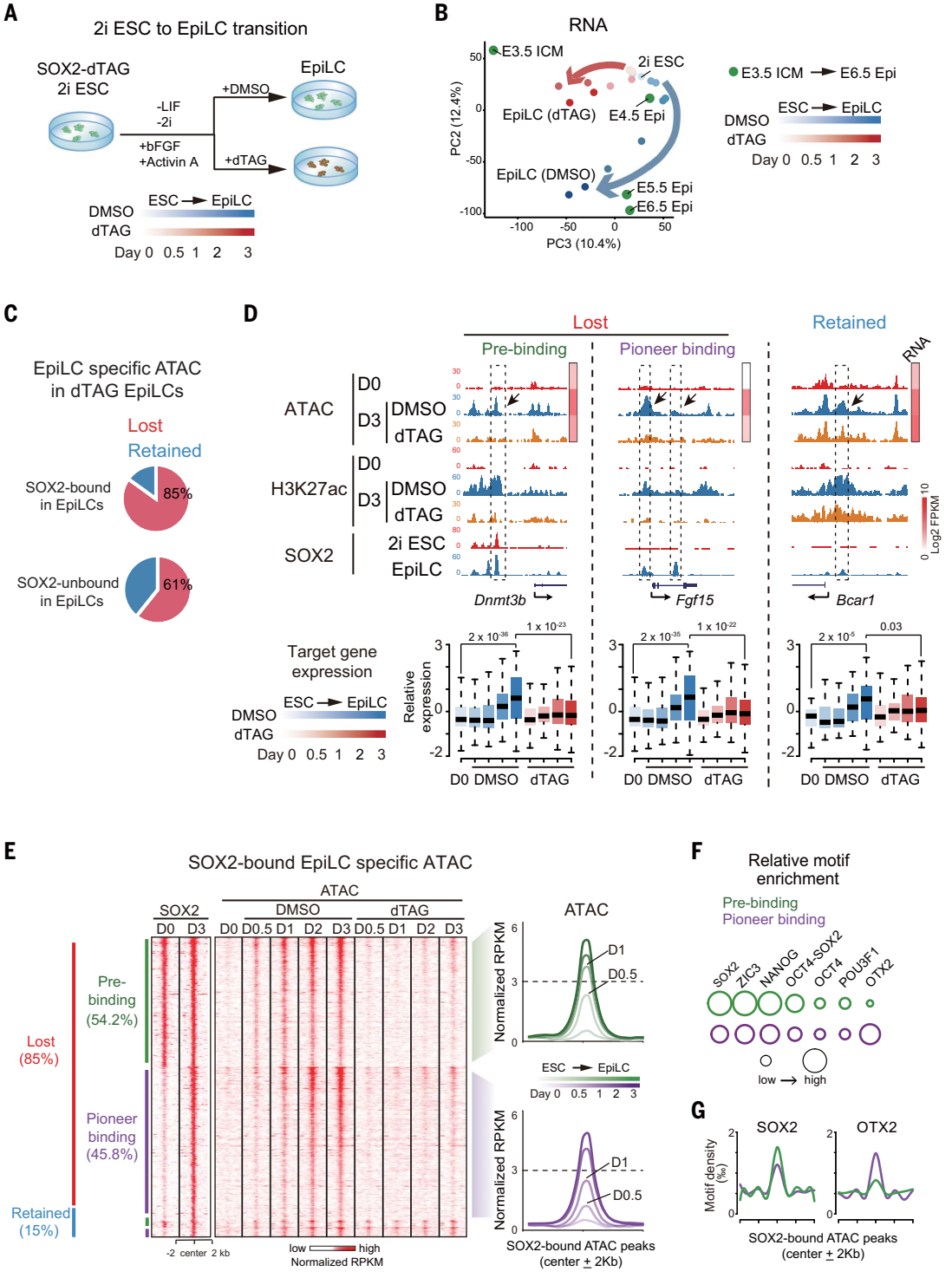


Fig. 6. Depletion of SOX2 impedes the naive-to-formative pluripotency transition. (**A**) Schematic showing differentiation of SOX2-dTAG 2i ESCs to EpiLCs with DMSO or dTAG treatment. (**B**) PCA showing RNA-seq of cells from 2i ESCs to EpiLCs at day 0 to day 3 with DMSO (blue) or dTAG (red) treatment. (**C**) Pie charts showing the percentages of lost and retained peaks at EpiLC-specific distal ATAC-seq peaks with or without SOX2 binding. (**D**) (Top) UCSC browser views and heatmaps showing SOX2 binding signals, ATAC-seq, H3K27ac enrichment, and gene expression of representative genes in 2i ESCs (D0) and DMSO- and dTAG-treated EpiLCs (D3). Arrows and dashed boxes indicate lost or retained ATAC-seq peaks. (Bottom) Box plots show the relative expression (row *z*-score normalized) of predicted target genes of peaks during the EpiLC transition, with *P* values (*t* test, two-sided) indicated on top. (**E**) (Left) Heatmaps showing enrichment of SOX2 binding and ATAC-seq signals at the lost and retained EpiLC-specific ATAC-seq peaks during the EpiLC transition with DMSO or dTAG treatment. The lost peaks are further clustered into SOX2 prebinding and pioneer binding sites in 2i ESCs. (Right) Line charts show the average ATAC-seq enrichment at the SOX2 prebinding or pioneer binding EpiLC-specific ATAC-seq peaks. (**F**) TF motifs identified from EpiLC-specific ATAC-seq peaks with SOX2 prebinding or pioneer binding. Sizes of circles indicate levels of −log *P* values. (**G**) Density plots showing corresponding TF motif density at SOX2 prebinding or pioneer binding EpiLC-specific ATAC-seq peaks.

pieces of evidence. (i) SOX2's binding peaks are less enriched for motifs of SOX2, OCT4, and OCT4-SOX2 but are more enriched for the motifs of early-stage TFs (such as NR5A2, TFAP2C, and GATA) (Fig. 1E). SOX2 binds pre-accessible chromatin in part opened by these early-stage TFs (Fig. 4E). (ii) The OCT4-SOX2 motif enrichment in SOX2-bound sites is strong in 2i ESCs and E4.5 epiblast but not in E3.5 ICM, which raises the possibility that the cooperative function of OCT4 and SOX2 may only become dominant after entering naive pluripotency. Accordingly, although both OCT4 and SOX2 promote ICM-specific genes and repress TE-specific genes in E3.5 ICM (*26*, *49*, *50*), their targets appear to differ (fig. S6B). Moreover, the transcriptional interdependence of master pluripotency TFs OCT4, SOX2, and NANOG in ESCs (*12*, *16*, *51*, *52*) was also not observed in E3.5 ICM. Unlike that in ESCs (*10*), *Sox2* KO in E3.5 ICM did not affect expression of *Oct4* or *Nanog* (Fig. 2C), and *Oct4* KO did not affect *Nanog* expression and only partially down-regulates *Sox2* (*26*, *53*, *54*) (fig. S6A). (iii) SOX2 is globally dispensable for enhancer opening in E3.5 ICM, whereas it is essential for opening enhancers genome-wide in 2i ESCs and E4.5

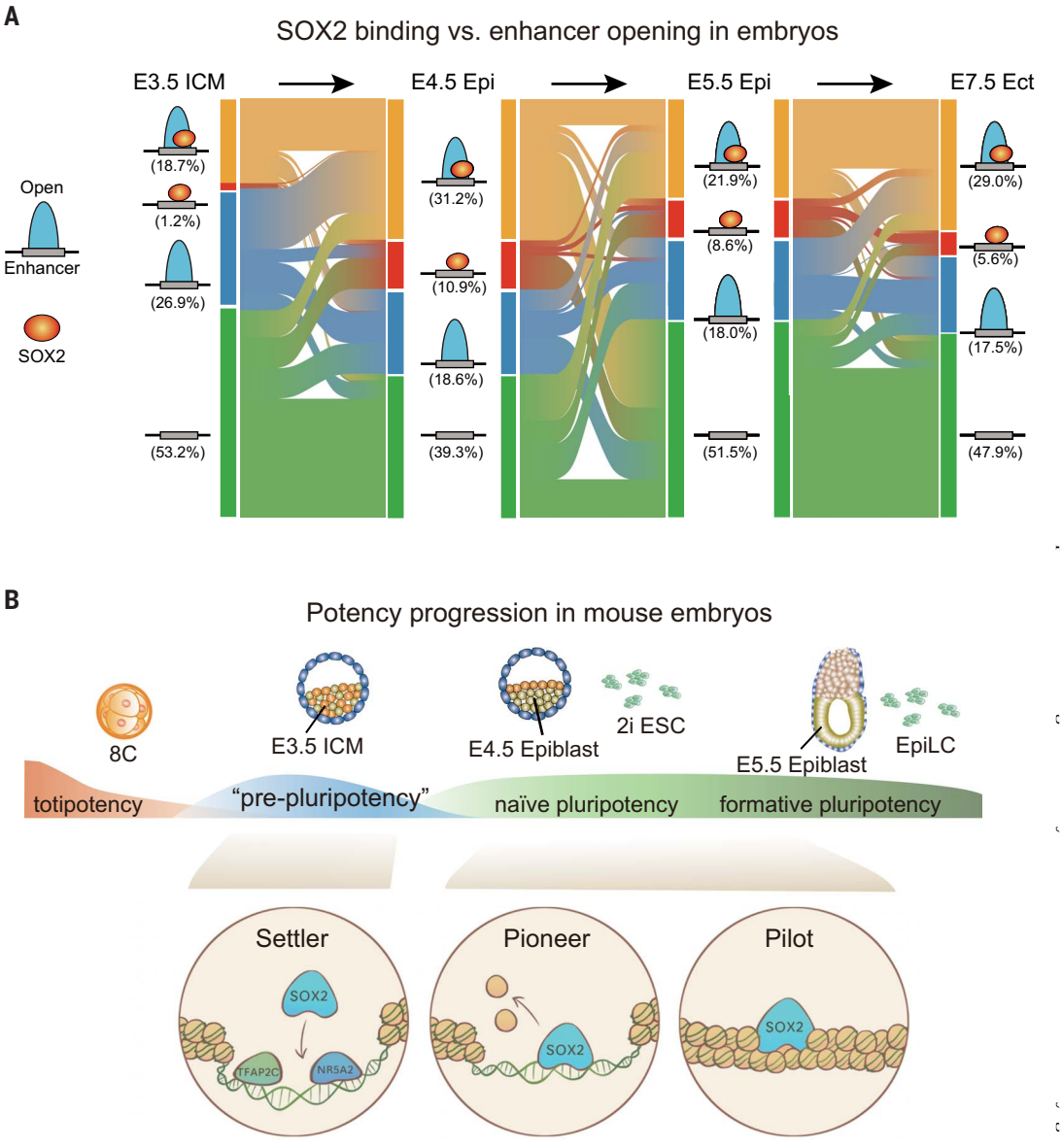


Fig. 7. SOX2-chromatin binding in early development and its multifaceted interaction modes with enhancers. (**A**) Alluvial diagrams showing the dynamics states of enhancers for their SOX2 binding and accessibility (based on ATAC-seq) in embryos. The percentages of each class of SOX2-chromatin interactions (combination of SOX2 occupancy and chromatin accessibility) within all enhancers (pooled from all stages examined) are shown for each stage. (**B**) A model illustrating the multifaceted SOX2 interaction modes with enhancers during the pluripotency transition. Three SOX2 binding modes are proposed: settler, pioneer, and pilot binding. The settler binding refers to SOX2 binding preaccessible enhancers, and its depletion does not substantially affect chromatin opening, as exemplified by most SOX2 binding in E3.5 ICM. The pioneer binding occurs in E4.5 epiblast or 2i mESCs, where SOX2 is required to establish or maintain the enhancer accessibility. The pilot binding of SOX2 at many formative enhancers in 2i mESCs is insufficient for enhancer opening but may help poise enhancers for faster opening upon conversion to formative pluripotency. E3.5 ICM is in a prepluripotency state to bridge totipotency to pluripotency, featured by coexpression of multilineage markers (epiblast, PrE, and early-stage TFs), expanded potency toward epiblast and PrE, and a primitive pluripotency network.

epiblast. A larger role of pluripotency factors in ESCs is consistent with them being required for mouse ESC derivation (*50*, *55*) and maintenance (*16*). Hence, the E3.5 ICM exhibits a distinct, "prepluripotency" state, featuring the potential to give rise to both epiblast and PrE, coexpression of multilineage TFs, and a primitive pluripotency network.

Pioneer TFs are believed to bind and open inaccessible chromatin, leading to the subsequent recruitment of additional TFs (*17*). We found that SOX2 manifests more diverse roles at enhancers beyond a simple pioneer factor, which include settler binding, pioneer binding, and pilot binding. SOX2 exhibits settler binding in E3.5 ICM, where SOX2 binds preaccessible enhancers, and its loss does not substantially affect chromatin opening (Fig. 7B). We speculate that the relatively short expression period and lack of cooperation with OCT4 or other cofactors may disable its pioneering binding function. Notably, such settler binding can still exert impacts on gene expression, especially at sites with SOX2 motifs (Fig. 3B). It is possible that the SOX2 motif may increase the residence time of SOX2, which in turn promotes gene expression, for example by increasing promoter-enhancer interactions (*8*, *56*). It also remains to be investigated whether some settler binding may help sequester excess SOX2 from other binding sites to prevent premature activation of later-stage genes. Widespread pioneer binding is then observed in E4.5 epiblast and 2i ESCs, where SOX2 is required for naive enhancers opening (Fig. 3D and Fig. 5C). Finally, the pilot binding of SOX2 at many formative enhancers in 2i ESCs is insufficient for enhancer opening but likely helps enhancers with weak formative TF motifs achieve faster opening upon differentiation (Fig. 6E). We propose that such multifaceted—rather than a universal pioneering—chromatin interacting modes may also hold true for other pioneer TFs to allow precise yet adaptable responses to developmental cues beyond pluripotency regulation.

Materials and methods summary

A detailed materials and methods section is provided in the supplementary materials. All animals were cared for according to the guidelines of the Institutional Animal Care and Use Committee of Tsinghua University. Embryos were collected from superovulated females crossed with males. To generate *Sox2* mzKO embryos, $Sox2^{flox/flox}$, Zp3-Cre females and $Sox2^{flox/flox}$, Stra8-Cre males were used for breeding. Immunosurgery was performed as reported previously (*57*) to remove TE and isolate ICM. ICMs were then incubated in TrypLE

and dissociated by repetitive pipetting using a Pasteur pipette. For scRNA-seq, individual E3.5 ICM cells were transferred into single-cell lysis buffer following the Smart-seq2 protocol, as described previously (*58*). E4.5 blastocysts were flushed from the uterus after human chorionic gonadotropin (hCG) injection at 114 to 116 hours. Given that SOX2 was present only in epiblast but not in PrE at E4.5 (*10*), we profiled SOX2 binding using the entire ICM because the signals were expected to arise exclusively from epiblast cells. E5.5 to E7.5 embryo tissues were collected as previously described (*23*, *59*, *60*).

CUT&RUN was conducted following the published protocol (*61*) with some modifications. The fresh samples were resuspended and bound with concanavalin-coated magnetic beads. After incubation with SOX2 antibody for 2 to 3 hours at 4°C, the samples were incubated with protein A–micrococcal nuclease (pA-MNase) for 1 hour. The STAR ChIP-seq for H3K27ac and miniATAC-seq were performed as previously described (*62*, *63*).

To construct SOX2-dTAG ESCs, the sequence encoding $FKBP^{F36V}$-GFP was fused to the C terminus of the endogenous *Sox2* locus. SOX2-FKBP proteins were depleted by adding $dTAG^{V}$-1 into the medium. Time-course experiments were performed by inducing protein degradation and collecting the samples at different time points. Naive mESCs (2i mESCs) were cultured in the N2B27 medium supplemented with PD0325901, Chir99021, and LIF. To induce EpiLC differentiation, 2i ESC cells were plated on tissue culture dishes pretreated with matrigel in N2B27-based medium supplemented with 1% knockout serum replacement (KSR), basic fibroblast growth factor (bFGF), and activin A.

REFERENCES AND NOTES

1. J. Rossant, P. P. L. Tam, New Insights into Early Human Development: Lessons for Stem Cell Derivation and Differentiation. *Cell Stem Cell* **20**, 18–28 (2017). doi: 10.1016/j.stem.2016.12.004; pmid: 28061351
2. M. Zernicka-Goetz, S. A. Morris, A. W. Bruce, Making a firm decision: Multifaceted regulation of cell fate in the early mouse embryo. *Nat. Rev. Genet.* **10**, 467–477 (2009). doi: 10.1038/nrg2564; pmid: 19536196
3. C. Chazaud, Y. Yamanaka, T. Pawson, J. Rossant, Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev. Cell* **10**, 615–624 (2006). doi: 10.1016/j.devcel.2006.02.020; pmid: 16678776
4. A. Smith, Formative pluripotency: The executive phase in a developmental continuum. *Development* **144**, 365–373 (2017). doi: 10.1242/dev.142679; pmid: 28143843
5. K. Hayashi, H. Ohta, K. Kurimoto, S. Aramaki, M. Saitou, Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011). doi: 10.1016/j.cell.2011.06.052; pmid: 21820164
6. J. Nichols, A. Smith, Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009). doi: 10.1016/j.stem.2009.05.015; pmid: 19497275
7. J. Rossant, Genetic Control of Early Cell Lineages in the Mammalian Embryo. *Annu. Rev. Genet.* **52**, 185–201 (2018). doi: 10.1146/annurev-genet-120116-024544; pmid: 30183407
8. F. Spitz, E. E. Furlong, Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012). doi: 10.1038/nrg3207; pmid: 22868264
9. R. A. Young, Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011). doi: 10.1016/j.cell.2011.01.032; pmid: 21414485
10. E. Wicklow *et al*., HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. *PLOS Genet.* **10**, e1004618 (2014). doi: 10.1371/journal.pgen.1004618; pmid: 25340657
11. A. A. Avilion *et al*., Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140 (2003). doi: 10.1101/gad.224503; pmid: 12514105
12. X. Chen *et al*., Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008). doi: 10.1016/j.cell.2008.04.043; pmid: 18555785
13. R. Blassberg *et al*., Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation. *Nat. Cell Biol.* **24**, 633–644 (2022). doi: 10.1038/s41556-022-00910-2; pmid: 35550614
14. S. J. Hainer, S. Bošković, K. N. McCannell, O. J. Rando, T. G. Fazzio, Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319–1329.e11 (2019). doi: 10.1016/j.cell.2019.03.014; pmid: 30955888
15. J. J. Thompson *et al*., Extensive co-binding and rapid redistribution of NANOG and GATA6 during emergence of divergent lineages. *Nat. Commun.* **13**, 4257 (2022). doi: 10.1038/s41467-022-31938-5; pmid: 35871075
16. S. Masui *et al*., Pluripotency governed by *Sox2* via regulation of *Oct3/4* expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635 (2007). doi: 10.1038/ncb1589; pmid: 17515932
17. K. S. Zaret, J. S. Carroll, Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011). doi: 10.1101/gad.176826.111; pmid: 22056668
18. A. Soufi *et al*., Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015). doi: 10.1016/j.cell.2015.03.017; pmid: 25892221
19. T. Boroviak, R. Loos, P. Bertone, A. Smith, J. Nichols, The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* **16**, 513–528 (2014). doi: 10.1038/ncb2965; pmid: 24859004
20. S. Okumura-Nakanishi, M. Saito, H. Niwa, F. Ishikawa, Oct-3/4 and Sox2 regulate *Oct-3/4* gene in embryonic stem cells. *J. Biol. Chem.* **280**, 5307–5317 (2005). doi: 10.1074/jbc.M410015200; pmid: 15557334
21. J. Wu *et al*., The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016). doi: 10.1038/nature18606; pmid: 27309802
22. B. Nabet *et al*., The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018). doi: 10.1038/s41589-018-0021-8; pmid: 29581585
23. Y. Xiang *et al*., Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. *Nat. Genet.* **52**, 95–105 (2020). doi: 10.1038/s41588-019-0545-1; pmid: 31844322
24. B. Plusa, A. Piliszek, S. Frankenberg, J. Artus, A. K. Hadjantonakis, Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**, 3081–3091 (2008). doi: 10.1242/dev.021519; pmid: 18725515
25. Y. Ohnishi *et al*., Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol.* **16**, 27–37 (2014). doi: 10.1038/ncb2881; pmid: 24292013
26. G. G. Stirparo *et al*., OCT4 induces embryonic pluripotency via STAT3 signaling and metabolic mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2008890118 (2021). doi: 10.1073/pnas.2008890118; pmid: 33452132
27. F. Lu *et al*., Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell* **165**, 1375–1388 (2016). doi: 10.1016/j.cell.2016.05.050; pmid: 27259149
28. R. I. Sherwood *et al*., Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014). doi: 10.1038/nbt.2798; pmid: 24441470
29. P. Gu *et al*., Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. *Mol. Cell. Biol.* **25**, 3492–3505 (2005). doi: 10.1128/MCB.25.9.3492-3505.2005; pmid: 15831456
30. U. Werling, H. Schorle, Transcription factor gene *AP-2γ* essential for early murine development. *Mol. Cell. Biol.* **22**, 3149–3156 (2002). doi: 10.1128/MCB.22.9.3149-3156.2002; pmid: 11940672
31. H. J. Auman *et al*., Transcription factor AP-2γ is essential in the extra-embryonic lineages for early postimplantation development. *Development* **129**, 2733–2747 (2002). doi: 10.1242/dev.129.11.2733; pmid: 12015300
32. Z. Cao *et al*., Transcription factor AP-2γ induces early *Cdx2* expression and represses HIPPO signaling to specify the trophectoderm lineage. *Development* **142**, 1606–1615 (2015). doi: 10.1242/dev.120238; pmid: 25858457
33. M. Zhu *et al*., Developmental clock and mechanism of de novo polarization of the mouse embryo. *Science* **370**, eabd2703 (2020). doi: 10.1126/science.abd2703; pmid: 33303584
34. J. Gassler *et al*., Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**, 1305–1315 (2022). doi: 10.1126/science.abn7478; pmid: 36423263
35. N. Festuccia, N. Owens, A. Chervova, A. Dubois, P. Navarro, The combined action of Esrrb and Nr5a2 is essential for murine naïve pluripotency. *Development* **148**, dev199604 (2021). doi: 10.1242/dev.199604; pmid: 34397088
36. N. Festuccia *et al*., *Nr5a2* is essential for morula development. bioRxiv 2023.01.16.524255 [Preprint] (2023); https://doi.org/10.1101/2023.01.16.524255.
37. F. Lai *et al*., NR5A2 connects zygotic genome activation to the first lineage segregation in totipotent embryos. *Cell Res.* 10.1038/s41422-023-00887-z (2023). doi: 10.1038/s41422-023-00887-z; pmid: 37935903
38. T. Boroviak *et al*., Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev. Cell* **35**, 366–382 (2015). doi: 10.1016/j.devcel.2015.10.011; pmid: 26555056
39. M. Kinoshita *et al*., Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. *Cell Stem Cell* **28**, 453–471.e8 (2021). doi: 10.1016/j.stem.2020.11.005; pmid: 33271069
40. D. Acampora, L. G. Di Giovannantonio, A. Simeone, Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development* **140**, 43–55 (2013). doi: 10.1242/dev.085290; pmid: 23154415
41. C. Buecker *et al*., Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014). doi: 10.1016/j.stem.2014.04.003; pmid: 24905168
42. S. H. Yang *et al*., ZIC3 Controls the Transition from Naive to Primed Pluripotency. *Cell Rep.* **27**, 3215–3227.e6 (2019). doi: 10.1016/j.celrep.2019.05.026; pmid: 31189106
43. F. Nakaki *et al*., Induction of mouse germ-cell fate by transcription factors *in vitro*. *Nature* **501**, 222–226 (2013). doi: 10.1038/nature12417; pmid: 23913270
44. J. Ernst *et al*., Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011). doi: 10.1038/nature09906; pmid: 21441907
45. M. Thomson *et al*., Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875–889 (2011). doi: 10.1016/j.cell.2011.05.017; pmid: 21663792
46. R. Argelaguet *et al*., Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019). doi: 10.1038/s41586-019-1825-8; pmid: 31827285
47. I. Muñoz-Sanjuán, A. H. Brivanlou, Neural induction, the default model and embryonic stem cells. *Nat. Rev. Neurosci.* **3**, 271–280 (2002). doi: 10.1038/nrn786; pmid: 11967557
48. P. Karagianni, P. Moulos, D. Schmidt, D. T. Odom, I. Talianidis, Bookmarking by Non-pioneer Transcription Factors during Liver Development Establishes Competence for Future Gene Activation. *Cell Rep.* **30**, 1319–1328.e6 (2020). doi: 10.1016/j.celrep.2020.01.006; pmid: 32023452
49. Y. I. Yeom *et al*., Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996). doi: 10.1242/dev.122.3.881; pmid: 8631266
50. J. Nichols *et al*., Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998). doi: 10.1016/S0092-8674(00)81769-9; pmid: 9814708
51. D. J. Rodda *et al*., Transcriptional regulation of nanog by OCT4 and SOX2. *J. Biol. Chem.* **280**, 24731–24737 (2005). doi: 10.1074/jbc.M502573200; pmid: 15860457
52. J. L. Chew *et al*., Reciprocal transcriptional regulation of *Pou5f1* and *Sox2* via the Oct4/Sox2 complex in embryonic stem cells. *Mol. Cell. Biol.* **25**, 6031–6046 (2005). doi: 10.1128/MCB.25.14.6031-6046.2005; pmid: 15988017

53. T. Frum *et al*., Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. *Dev. Cell* **25**, 610–622 (2013). doi: 10.1016/j.devcel.2013.05.004; pmid: 23747191
54. G. C. Le Bin *et al*., Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. *Development* **141**, 1001–1010 (2014). doi: 10.1242/dev.096875; pmid: 24504341
55. K. Mitsui *et al*., The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003). doi: 10.1016/S0092-8674(03)00393-3; pmid: 12787504
56. R. Stadhouders, G. J. Filion, T. Graf, Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345–354 (2019). doi: 10.1038/s41586-019-1182-7; pmid: 31092938
57. D. Solter, B. B. Knowles, Immunosurgery of mouse blastocyst. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 5099–5102 (1975). doi: 10.1073/pnas.72.12.5099; pmid: 1108013
58. S. Picelli *et al*., Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014). doi: 10.1038/nprot.2014.006; pmid: 24385147
59. S. M. Harrison, S. L. Dunwoodie, R. M. Arkell, H. Lehrach, R. S. Beddington, Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissueconstituents of the gastrulating mouse embryo. *Development* **121**, 2479–2489 (1995). doi: 10.1242/dev.121.8.2479; pmid: 7671812
60. Y. Zhang *et al*., Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat. Genet.* **50**, 96–105 (2018). doi: 10.1038/s41588-017-0003-x; pmid: 29203909
61. P. J. Skene, J. G. Henikoff, S. Henikoff, Targeted *in situ* genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018). doi: 10.1038/nprot.2018.015; pmid: 29651053
62. B. Zhang *et al*., Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016). doi: 10.1038/nature19361; pmid: 27626382
63. J. Wu *et al*., Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018). doi: 10.1038/s41586-018-0080-8; pmid: 29720659

ACKNOWLEDGMENTS

We are grateful to members of the Xie laboratory for discussions and comments during the SOX2 study and the preparation of the manuscript and to the Animal Center and Biocomputing Facility at Tsinghua University for their support. We are grateful to Q. Sun from the Institute of Zoology, Chinese Academic of Sciences, for providing us with the Stra8-cre mice. **Funding:** This work was funded by the National Key R&D Program of China (2021YFA1100102 to W.X.); the National Natural Science Foundation of China (31988101 and 31830047 to W.X.); the National Key R&D Program of China (2019YFA0508900 to W.X.); and the Tsinghua-Peking Center for Life Sciences (to W.X.) M.A.H., T.F., and A.R. have been supported by NIH R35 GM131759 and R01 HD108722. W.X. is a recipient of an HHMI International Research Scholar and a New Cornerstone Investigator. **Author contributions:** L.Li, F.L., and W.X. conceived and designed the project. L.Li performed the CUT&RUN and RNA-seq experiments. F.L. performed embryo experiments with help from X.L., Z.L., L.Liu, and Y.X. F.L. performed microinjection and immunostaining with the help of L.Liu. L.Li performed dTAG and EpiLC differentiation experiments. X.H. performed STAR ChIP-seq and constructed plasmids for reporter assay and rescue experiments. B.L. and L.Li performed ATAC-seq experiments. L.Li performed the bioinformatics analysis with the help of B.L. and F.C. L.Li and W.X. prepared most figures and wrote the manuscript with help from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The datasets generated and analyzed during this study are available in the Gene Expression Omnibus (GEO) database under accession no. GSE203194. Accession codes of the published data in GEO used in this study are as follows: RNA-seq and ATAC-seq of early embryos, GSE66390; RNA-seq and ATAC-seq of postimplantation epiblast and ectoderm, GSE125318; H3K27me3 of 2-cell, GSE76687; scRNA-seq of E4.5 epiblast and PrE, GSE159030; morula DNase-seq data, GSE92605; SOX2/NANOG ChIP of serum mESC, GSM2417143; H3K4me1 and OCT4/OTX2 ChIP of mESC and EpiLCs, GSM1355167; Hi-C of ESCs, GSE118911; Hi-C of EpiLCs, GSE183828; scRNA of Oct4 KO ICM, GSE159030; NR5A2 CUT&RUN and RNA-seq in the *Nr5a2* knockdown embryos, GSE229740; and TFAP2C CUT&RUN and RNA-seq in the *Tfap2c* knockout mouse embryos, GSE203194. Accession code of the published data in ArrayExpress is as follows: ZIC3 ChIP-seq data in mESCs and EpiLCs, E-MTAB-7208.
License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi5516
Materials and Methods
Figs. S1 to S14
References (*64*–*77*)
MDAR Reproducibility Checklist

Submitted 4 May 2023; accepted 9 November 2023
10.1126/science.adi5516

RESEARCH ARTICLE SUMMARY

MICROBIOTA

Microbiome diversity protects against pathogens by nutrient blocking

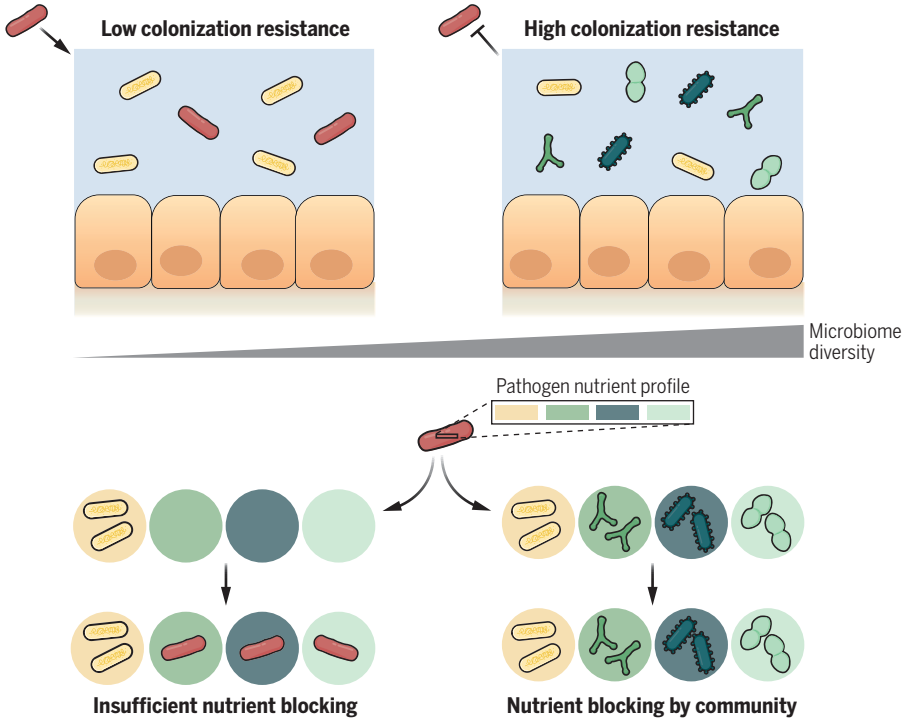
Frances Spragge†, Erik Bakkeren†, Martin T. Jahn, Elizete B. N. Araujo, Claire F. Pearson, Xuedan Wang, Louise Pankhurst, Olivier Cunrath*, Kevin R. Foster*

INTRODUCTION: The diverse bacterial species that colonize the human gut, which are collectively known as the gut microbiota, provide important health benefits. One of the key benefits is colonization resistance—the ability to restrict colonization of the gut by pathogens that can trigger disease. Multiple mechanisms have been found to influence the ability of the microbiota to provide colonization resistance, but these mechanisms are often context-specific and dependent on particular strains or species of bacteria. As a result, we lack general principles to predict which microbiota communities will be protective versus those that will allow pathogens to colonize.

RATIONALE: We used an ecological approach to study the colonization resistance provided by human gut symbionts against two important bacterial pathogens, *Klebsiella pneumoniae* and *Salmonella enterica* serovar Typhimurium. We studied colonization resistance provided by symbionts both alone and in combinations of increasing diversity to identify general patterns underlying colonization resistance, using both in vitro assays and in vivo work with gnotobiotic mice.

RESULTS: We cultured 100 human gut symbionts individually with *K. pneumoniae* and then *S.* Typhimurium and ranked the symbionts on the basis of their ability to provide colonization resistance. However, even the best-performing species provided limited protection against the pathogens in our assays. By contrast, when we combined species into diverse communities of up to 50 species, we found cases in which pathogen growth was greatly limited. The same patterns were observed w germ-free mice were colonized by a subse these communities and challenged with a pathogen. Ecological diversity, therefore, was important for colonization resistance, but we also found that community composition was important. Both in vitro and in vivo, we found that colonization resistance rested upon certain species being present, even though these species offer little protection on their own. We were able to explain these patterns from the ability of some communities to block pathogen growth by consuming the nutrients that the pathogen needs. Nutrient blocking is thus promoted both by diversity and by the presence of certain key species that increase the overlap between the nutrient use of a community and a pathogen. As a result, the inclusion of a key species closely related to a pathogen can be central to making a community protective because it provides a higher degree of metabolic overlap. However, this alone is typically not sufficient. We found that the presence of additional, often distantly related species is also needed to ensure that nutrient blocking—and consequently, colonization resistance—occurs. Lastly, we used the nutrient-blocking principle to predict in silico more-protective and less-protective communities for a new target strain, an antimicrobial resistant *Escherichia coli* clinical isolate. We then tested the colonization resistance of these communities experimentally. This work revealed that we can successfully identify protective communities from a large number of possible combinations, using both phenotypic measures of metabolic overlap but also a more general measure of genomic overlap.

CONCLUSION: Our results support the idea that more-diverse microbiomes can provide health benefits, specifically that they can improve protection against pathogen colonization. We also find that colonization resistance is a collective property of microbiome communities; in other words, a single strain is protective only when in combination with others. Crucially, although increased microbiome diversity increases the probability of protection against pathogens, the overlap in nutrient-utilization profiles between the community and the pathogen is key. Our work suggests a route to optimize the composition of microbiomes for protection against pathogens. ■



Microbiome diversity protects against pathogens by nutrient blocking. Pathogens (red) fail to colonize when they overlap with the community (yellow and green bacteria) in nutrient-utilization profiles (nutrient niches are indicated by colored circles). As microbiome diversity increases, the probability that different nutrients are consumed increases, which helps to block pathogen growth and improve colonization resistance.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: kevin.foster@biology.ox.ac.uk (K.R.F.); olivier.cunrath@unistra.fr (O.C.)
†These authors contributed equally to this work.
Cite this article as F. Spragge *et al.*, *Science* **382**, eadj3502 (2023). DOI: 10.1126/science.adj3502

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adj3502

RESEARCH ARTICLE

MICROBIOTA

Microbiome diversity protects against pathogens by nutrient blocking

Frances Spragge[1,2]†, Erik Bakkeren[1,2]†, Martin T. Jahn[1,2], Elizete B. N. Araujo[3], Claire F. Pearson[3], Xuedan Wang[1,2], Louise Pankhurst[1,2], Olivier Cunrath[4]*, Kevin R. Foster[1,2]*

The human gut microbiome plays an important role in resisting colonization of the host by pathogens, but we lack the ability to predict which communities will be protective. We studied how human gut bacteria influence colonization of two major bacterial pathogens, both in vitro and in gnotobiotic mice. Whereas single species alone had negligible effects, colonization resistance greatly increased with community diversity. Moreover, this community-level resistance rested critically upon certain species being present. We explained these ecological patterns through the collective ability of resistant communities to consume nutrients that overlap with those used by the pathogen. Furthermore, we applied our findings to successfully predict communities that resist a novel target strain. Our work provides a reason why microbiome diversity is beneficial and suggests a route for the rational design of pathogen-resistant communities.

The human gut is home to diverse bacterial species collectively known as the gut microbiota. A major health benefit provided by the gut microbiota is protection against pathogen colonization and subsequent infection; a phenomenon known as colonization resistance (*1*). The ability of the microbiota to protect against numerous enteric pathogens is well documented, with evidence that particular species within the microbiota play a more important role than others (*2*–*9*). The ways that colonization resistance can arise include competition for nutrients and space, direct antagonism by toxins and other harmful compounds, and promoting host immunity against pathogens (*1*, *10*, *11*).

However, although the importance of the microbiota for colonization resistance is clear, we currently lack the principles needed to predict, a priori, which microbiota species will be effective against a given pathogen. A key challenge is the ecological complexity of the gut. The gut microbiome is a diverse ecological system with many individual species that all have the potential to play a role in colonization resistance. Moreover, these constituent species can also affect each other and interact ecologically in ways that are critical for colonization resistance (*12*–*16*). This combination of species diversity and the potential for ecological interactions makes colonization resistance a challenging phenotype to understand (*17*).

We approached the question of mechanisms of colonization resistance from the perspective of the underlying ecological principles. To do this, we studied colonization resistance provided by a range of human gut bacteria, both alone and in combinations. We performed all experiments in parallel using two species of pathogen, which are both on the World Health Organization priority list: *Klebsiella pneumoniae* and *Salmonella enterica* serovar Typhimurium (*18*). Both are members of the *Enterobacteriaceae* found in the human gut microbiome, but they have very different lifestyles. *S.* Typhimurium causes acute infection and gastroenteritis (*19*, *20*). By contrast, *K. pneumoniae* is a nosocomial, opportunistic pathogen that rarely causes disease in the gut itself, but gut colonization with *K. pneumoniae* is a major risk factor for antimicrobial resistance–associated infections elsewhere in the body (*21*).

Despite these differences, we have identified common principles that underlie colonization resistance to both species. Ecological diversity is important for colonization resistance in vitro and in gnotobiotic mice. Moreover, we found that colonization resistance is an ecologically complex trait, whereby the protection against pathogens provided by one species can increase greatly in the presence of other species (*22*). Despite this complexity, we find that these ecological patterns are explained by a simple underlying principle: the collective ability of certain communities to consume nutrients and block pathogen growth. Furthermore, we have shown that this principle offers a way to identify sets of bacterial species that will collectively limit the growth of a particular pathogen.

Results

Single species offer little protection in competition with pathogens

Individual members of the microbiota can promote colonization resistance in various contexts (*2*–*8*), which suggests that some species are more important for colonization resistance than others. To systematically assess this variability, we screened a diverse set of 100 human gut symbionts (table S1 and Materials and methods) for their ability to limit pathogen growth. Competition in the gut occurs both at the point when a pathogen enters the gut and when a pathogen becomes established (*23*, *24*). We designed two coculture assays to reflect these two aspects of competition in the mammalian gut (Fig. 1A). In the first assay (ecological invasion assay), we pre-grew the symbiont alone in standard anaerobic media [modified Gifu anaerobic media (mGAM)] buffered to human colonic pH before adding the pathogen. In the second assay (competition assay), we inoculated this media with an equal ratio of symbiont to pathogen, which is an approach designed to capture competition once a pathogen has established itself in the gut.

To assess pathogen growth, we built luminescent strains of *K. pneumoniae* and *S.* Typhimurium and compared luminescence when grown in monoculture and when grown in coculture with each symbiont. With this assay system, we could rank the strains based on their abilities to limit pathogen growth in both the ecological invasion and competition assays (Fig. 1, B and C, and fig. S1). From this ranking, we took the top 10 best-performing nonpathogenic symbiont species in the screen [Materials and methods; fig. S1, E and F (orange circles); and table S1] and subjected them to a more stringent test of colonization resistance designed to capture both phases of competition in the gut in one assay (extended competition assay) (Fig. 1D). In this assay, the pathogen is first introduced into a pre-grown culture of a given symbiont strain, and after 24 hours, the mixture is passaged into fresh media and allowed to grow for 24 hours, at which time pathogen abundance is assessed with flow cytometry (Fig. 1D). Despite choosing the best-ranked species from the luminescence screen, all symbionts performed poorly under extended competition, with the majority offering no discernible colonization resistance (Fig. 1, E and F). The best performer was *Escherichia coli*, a known competitor of *S.* Typhimurium and also a member of the *Enterobacteriaceae*, but even here the protection offered was very limited, with the pathogens still able to reach 10^8 to 10^9 cells/ml.

The outcome of the assay differed greatly when we pooled all 10 species together (Fig. 1, E and F). Now, the final abundance of both pathogens was strongly suppressed by more than three orders of magnitude for *K. pneumoniae* and about two orders of magnitude for *S.* Typhimurium. By contrast, a community made up of the 10 worst-performing species from the luminescence screen (fig. S1, E and F, blue circles) provided little or no colonization resistance (Fig. 1, E and F). These results, therefore,

[1]Department of Biology, University of Oxford, Oxford, UK. [2]Department of Biochemistry, University of Oxford, Oxford, UK. [3]Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK. [4]CNRS, UMR7242, Biotechnology and Cell Signaling, University of Strasbourg, Illkirch, France.
*Corresponding author. Email: kevin.foster@biology.ox.ac.uk (K.R.F.); olivier.cunrath@unistra.fr (O.C.)
†These authors contributed equally to this work.

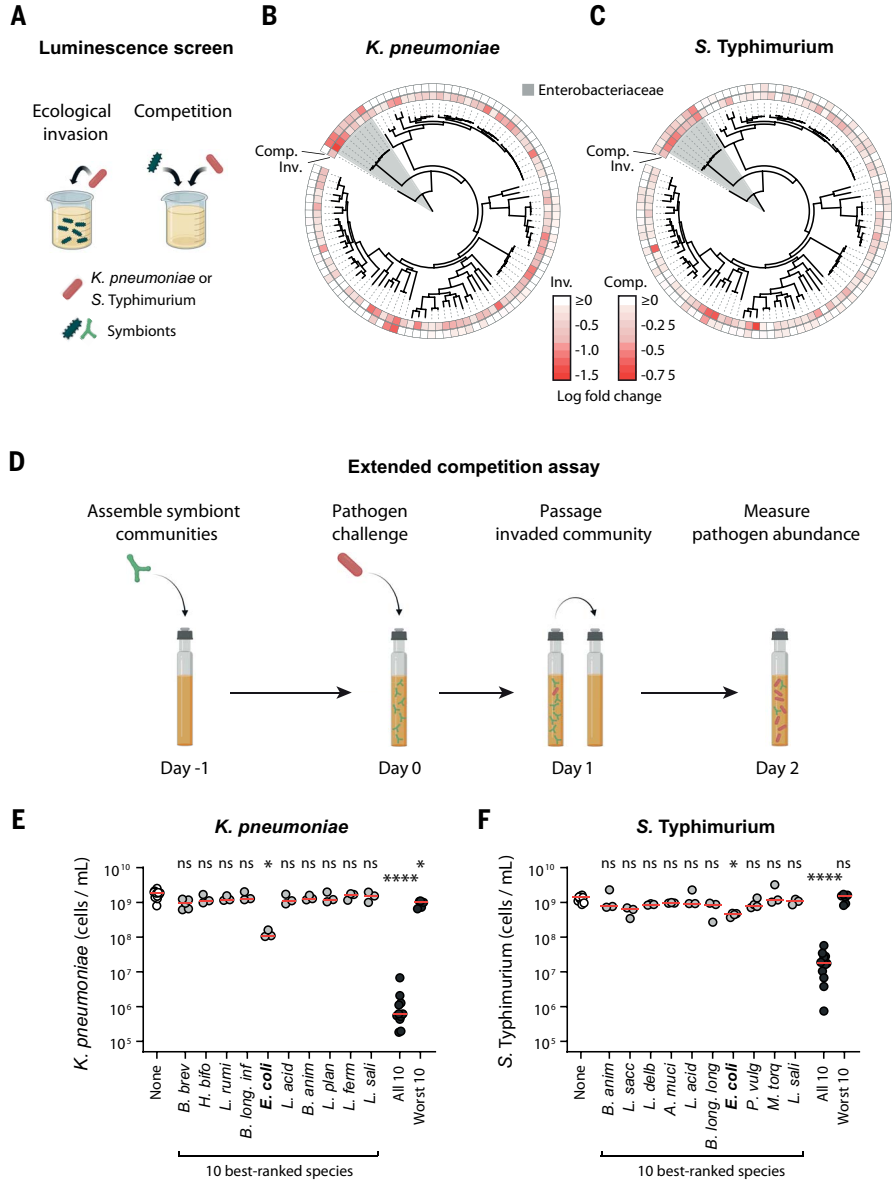


Fig. 1. Single species do not provide robust colonization resistance, but a diverse community can, depending on its composition. (**A**) Overview of the luminescence coculture assays. In the ecological invasion assay, *K. pneumoniae* or *S.* Typhimurium (red) was inoculated in coculture with individual symbionts (the two different green symbols in the key are used to represent the diversity of symbiont strains screened; 19:1 ratio of symbiont to pathogen). In the competition assay, the symbionts were inoculated at an equal ratio to the pathogen to recapitulate competition between strains once a pathogen is established in the microbiome. In both assays, luminescence produced by the pathogen was used as a proxy for pathogen growth. Created with BioRender.com. (**B** and **C**) Comparison of phylogenetic relatedness between symbionts and the ability of each symbiont to compete with the pathogen (Inv., ecological invasion assay; Comp., competition assay). Data for *K. pneumoniae* are shown in (B); data for *S.* Typhimurium are shown in (C). The family *Enterobacteriaceae*, which includes both *K. pneumoniae* and *S.* Typhimurium, is shaded in gray. Luminescence fold-change values are presented in fig. S1. Data are presented as the median luminescence log fold change of $n = 3$ to 10 independent experiments (biological replicates). Strains with the most negative (most red) values inhibited growth of the pathogen most strongly. (**D**) Overview of the extended competition assay. Communities (or individual strains; green) of symbionts were pre-grown in anaerobic-rich media before addition of the pathogen (red). The community was passaged after 24 hours of growth, followed by another 24 hours of growth before quantification with flow cytometry. Created with BioRender.com. (**E** and **F**) The extended competition assay was performed for each individual species identified in the 10 best-ranked species, as well as for combinations of 10 species (of both the 10 best- and worst-ranked species; fig. S1). Individual biological replicates from $n = 3$ to 15 independent experiments are shown. Red lines indicate the median. A Kruskal-Wallis test with Dunn's multiple test correction compares each group with the no-symbionts control [$P > 0.05$ = not significant (ns); *$P < 0.05$; ****$P < 0.0001$]. Data for *K. pneumoniae* are shown in (E); data for *S.* Typhimurium are shown in (F). See table S1 for species-name abbreviations.

suggest that strain identity is important for colonization resistance only in the context of a diverse community.

Ecological diversity and complexity drive colonization resistance in vitro

Our results indicated that microbiota diversity is important for colonization resistance. This finding fits well with the general idea that microbial diversity is beneficial for microbiome functioning, whereas a loss of diversity, or dysbiosis, can be associated with poor health and disease (*25*–*27*). Although the potential benefits of diversity are clear, cause and effect can be confounded in observational studies (*28*). To systematically test the role of diversity in colonization resistance, we randomly selected communities of increasing diversity from the 10 best-ranked species and competed them against the pathogens in the extended competition assay. To further evaluate the importance of diversity, we also assembled a community of 50 nonpathogenic symbiont species from the strains in our initial luminescence screen (Materials and methods). These data indicated a relationship between diversity and colonization resistance. However, we also saw a large variation in colonization resistance across the communities that differed in their composition of two, three, and five species. Visual inspection of the data (Fig. 2, C and D) suggested that a large component of this variability was driven by the composition of the communities.

One species that appeared to be important for outcomes was *E. coli*. To explore this finding, we randomly selected additional *E. coli*–containing communities and again evaluated colonization resistance (Fig. 2, C and D). We also performed dropout experiments in which we made up the 10- and 50-species communities without *E. coli* (Fig. 2, C and D). These data revealed a strong and clear monotonic increase in colonization resistance as species diversity increased (Fig. 2, C and D, green circles), but this relationship was much weaker or disappeared entirely in the absence of *E. coli* (fig. S2, A and B). In ecological terms, these data show that colonization resistance rests upon a strong higher-order effect involving other community members and *E. coli* (*22*). By higher-order effects, in this study we mean cases in which the effect of one species on another is changed by the presence of a third-party species in a community (*22*). That is, whereas either *E. coli* alone or the rest of the community without *E. coli* has little impact on pathogen growth, together they have a strong effect on pathogen growth. Such higher-order effects are considered important in ecology because they imply context dependence, which can make a system difficult to understand and predict (*22*, *29*–*32*). Another way to illustrate the effect of diversity on colonization resistance is to compare our data with a simple null

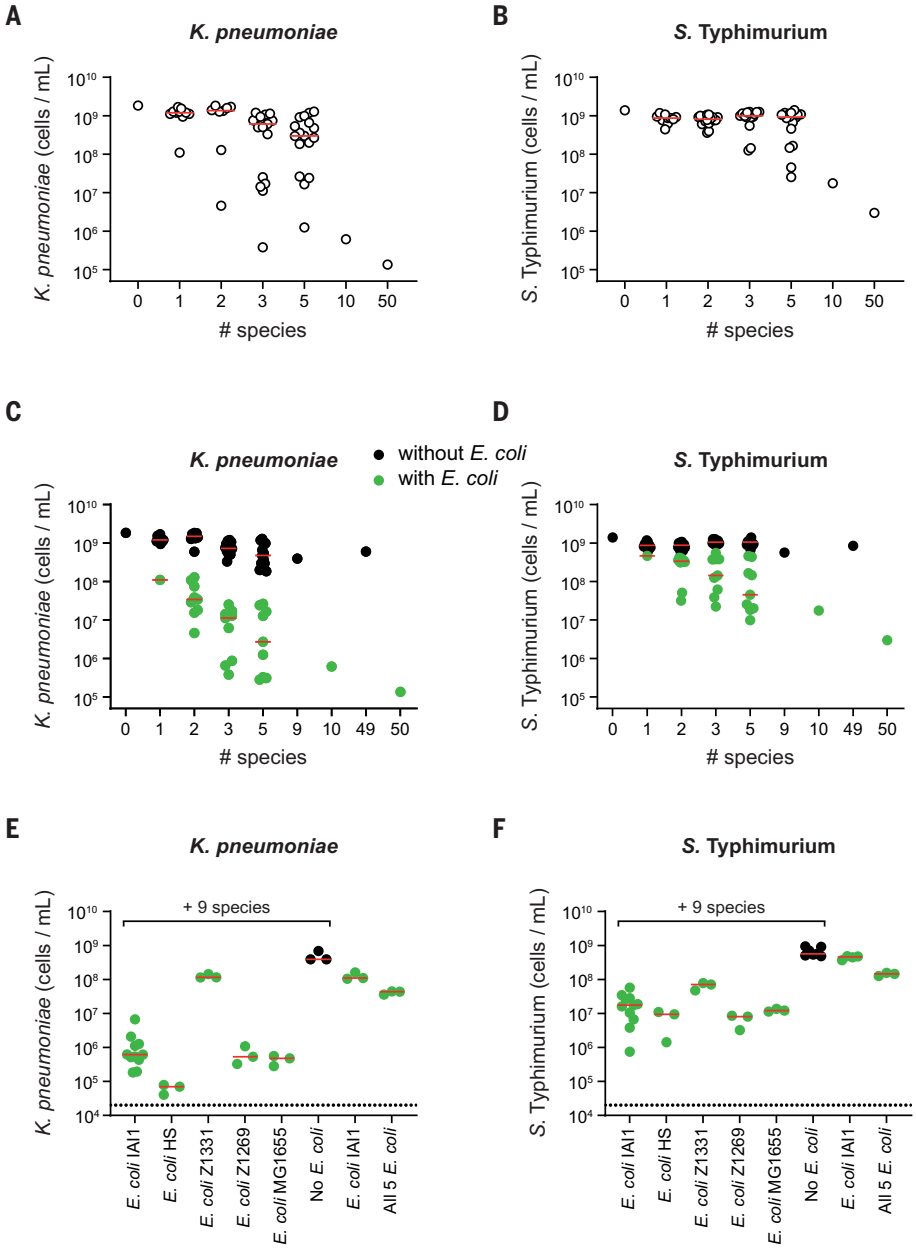


Fig. 2. Ecological diversity and key members are needed for efficient colonization resistance in vitro. (**A** to **D**) Extended competition assay on communities made up of an increasing number of species. Each data point represents the median pathogen cells/ml value on day 2 of the extended competition for a community (n = 3 to 15 biological replicates from independent experiments for each community; up to 17 communities for each group). Communities with size ≤10 species were randomly selected from the 10 best-ranked species for each pathogen. Community identities are shown in tables S4 and S5. Data for *K. pneumoniae* are shown in (A) and (C); data for *S.* Typhimurium are shown in (B) and (D). Red lines indicate the median value of communities at a given diversity level. In (C) and (D), data from (A) and (B) are replotted along with additional communities that always contained *E. coli* but were otherwise randomly selected. Communities without *E. coli* are depicted in black; communities with *E. coli* are depicted in green. Separate red median lines are shown for communities with and without *E. coli*. A linear regression is performed on log-log transformed data in fig. S2, A and B, which shows that the association between diversity and colonization observed is statistically significant and that this effect is greater for communities with *E. coli* than those without (F tests, $P \leq 0.0001$). (**E** and **F**) Results of extended competition assay testing *E. coli* strains substituted into the community of the 10 best-ranked species. Data for *K. pneumoniae* are shown in (E); data for *S.* Typhimurium are shown in (F). Red lines indicate median values. Each data point represents a biological replicate from independent experiments (n = 3 to 11).

model. Consider, for example, a model in which each additional species proportionally improves colonization resistance to the pathogen. We can compare our experimental data for *E. coli*–containing communities to a null model in which pathogen abundance scales according to $1/n$, where n is the number of species present. This analysis shows that the deviation from such a null model increases as diversity increases, where colonization resistance is again greater than expected for diverse communities (fig. S3).

We also asked whether the role of *E. coli* within communities was a strain-specific effect. We replaced *E. coli* strain IAI1, identified in our screen, with each of four other *E. coli* strains historically isolated from the human gut (*33*–*35*). The effect of *E. coli* was similar when *E. coli* IAI1 was substituted with most *E. coli* strains (Fig. 2, E and F), which indicates that the higher-order effect involving *E. coli* is a general property of closely related strains.

Further inspection of the data pointed to other species that were important for colonization resistance in diverse communities. In *E. coli*–containing communities, the presence of *Bifidobacterium breve* appeared to be important in excluding *K. pneumoniae* (fig. S4A), and the presence of *Lacrimispora saccharolyticum* and *Phocaeicola vulgatus* appeared important in the exclusion of *S.* Typhimurium (fig. S4B). We confirmed these patterns through a series of systematic dropout experiments (fig. S4, C and D). However, it was still possible to achieve equivalent colonization resistance in more diverse communities that lack these species (fig. S4, C and D), which again points to the underlying benefits of a diverse microbiota.

Ecological diversity and complexity also drive colonization resistance in vivo

To validate our in vitro methods, we tested the ability of symbiont communities to resist pathogen colonization in gnotobiotic mice (Fig. 3A). Germ-free mice were colonized with symbiont communities differing in diversity and in the presence or absence of *E. coli*. Successful colonization by *S.* Typhimurium causes an acute infection and massive gut inflammation, which is a major confounding effect for studying the effects of community composition on pathogen growth. Animals with a less-protective microbiota can rapidly succumb to the infection, such that one cannot follow ecological dynamics over time in a comparable way across treatments. We thus chose to use an avirulent variant of *S.* Typhimurium to eliminate the effect of gut inflammation on pathogen and host, allowing pathogen abundance to be used as a measure of disease risk (*19*, *36*, *37*). We introduced communities across the same range of diversities as before; in contrast to in vitro assays, however, not all symbiont species will reliably colonize germ-free mice (*38*). Therefore,

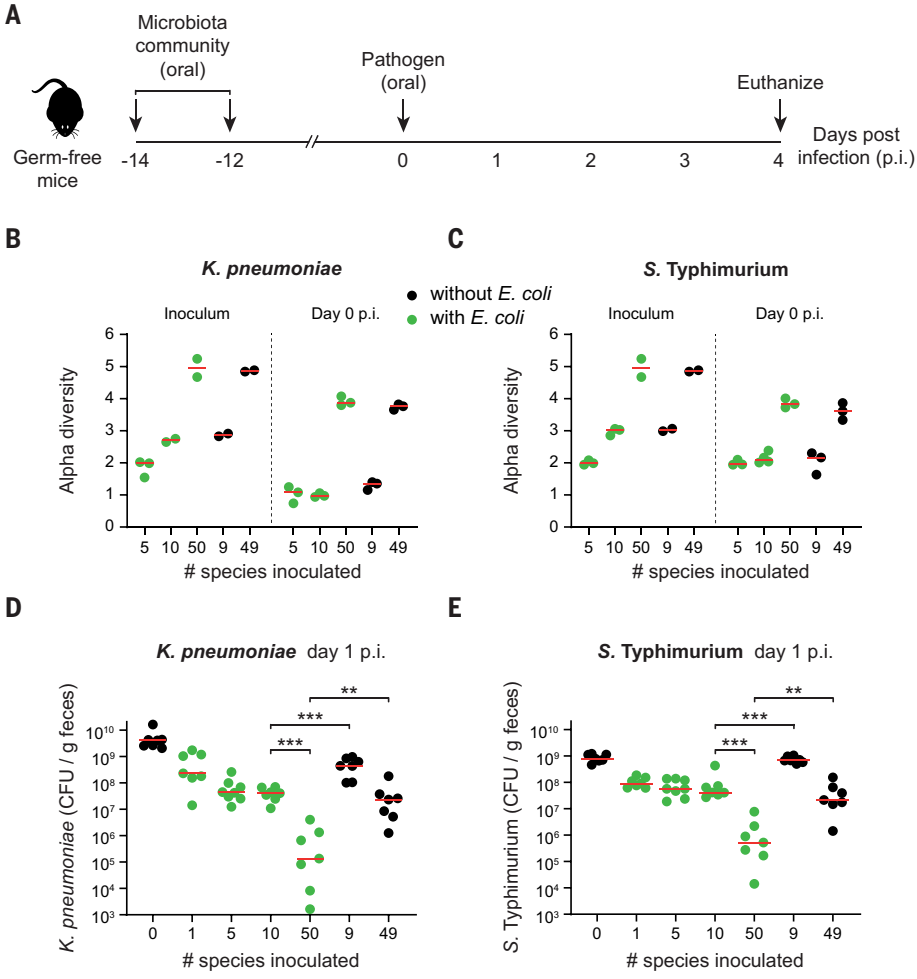


Fig. 3. Ecological diversity and key members are needed for efficient colonization resistance in vivo. (**A**) Overview of gnotobiotic mouse experiments. Symbiont communities (or *E. coli* alone) were given to germ-free mice by oral gavage twice (2 days apart). Twelve days later, the mice were challenged with *K. pneumoniae* or *S.* Typhimurium by oral gavage. Feces were collected from mice daily before mice were euthanized on day 4 postinfection (p.i.). (**B** and **C**) Alpha diversity measured with Shannon index of symbiont communities. Metagenomic sequencing was performed on the inoculum and fecal samples at day 0 (when the pathogen was introduced) and used to calculate diversity. Data for *K. pneumoniae* are shown in (B); data for *S.* Typhimurium are shown in (C). Biological replicates from a representative mouse from each cage are shown (n = 2 to 4; at least two independent experiments). (**D** and **E**) Pathogen abundances in the feces of gnotobiotic mice colonized with communities of increasing diversity (mice containing communities with *E. coli* shown in green; mice containing communities without *E. coli* shown in black; n = 7 to 8 biological replicates of mice per group in cages of two to three mice; two to three independent experiments). Red lines indicate the medians. Two-tailed Mann-Whitney tests are used to compare the indicated groups (**P < 0.01; *** P < 0.001). Data for *K. pneumoniae* are shown in (D); data for *S.* Typhimurium are shown in (E). Metagenomic analysis of species diversity and relative abundance is shown in fig. S5. Pathogen abundance data from days 1 to 4 p.i. are shown in fig. S6. Community compositions are shown in table S6.

we used metagenomic sequencing to confirm that introducing a higher-diversity community to the mice did indeed result in a higher diversity of species colonizing the gut—as measured by two metrics of alpha diversity—and to identify the relative abundance of all members (Fig. 3, B and C, and fig. S5).

These experiments revealed that, as observed in vitro, microbiome diversity is negatively correlated with pathogen abundance in feces for both pathogens (compare 10- versus 50-member communities) (Fig. 3, D and E, and fig. S6). Moreover, dropout experiments again revealed the importance of the combination of *E. coli* and other community members for colonization resistance (Fig. 3, D and E). We also observed that higher diversities are needed for efficient colonization resistance in the mammalian gut than in our in vitro assays, which is likely to be explained by the higher degree of environmental and spatial heterogeneity in the gut compared with that found in a test tube. Nevertheless, the key patterns remained the same between the gnotobiotic mouse experiments and our in vitro assays. Both ecological diversity and higher-order interactions were important for colonization resistance to both pathogens. As before, we saw a strong deviation from a simple null model of ecological competition at high levels of diversity (fig. S3). In addition to showing the generality of these patterns, this fit between the in vitro and in vivo methods validates our extended competition assay as an approach to interrogate the ecology of colonization resistance.

A simple principle explains the roles of diversity and complexity in colonization resistance

The discovery of such higher-order effects in colonization resistance indicates that colonization resistance is an ecologically complex trait (*22*), one which can be challenging to work with owing to high levels of context dependence (*17*, *22*, *29*, *30*). Nevertheless, we sought to understand the mechanisms underpinning colonization resistance by returning to our in vitro data gathered from large numbers of different communities. We used the species' genomes to assess functional similarity between symbiont communities and pathogens from overlap in protein compositions. Specifically, we calculated the percentage of all protein families carried by a pathogen that were also present in each community investigated (Materials and methods). We reasoned that this measure of functional similarity may map to niche overlap and, therefore, to the strength of ecological competition between symbionts and pathogens. We first confirmed that the number of encoded protein families covered by our experimental communities increases proportionally with the number of added species (fig. S7). Permutation analyses also confirmed that the randomly selected communities we have studied experimentally are a good representation of all possible communities that we could have studied (fig. S8).

The potential importance of protein-family overlap was already clear from the effects of *E. coli* in our experimental data (Figs. 2 and 3). *E. coli* is in the same family of bacteria as *K. pneumoniae* and *S.* Typhimurium and can be seen to contribute greatly to the protein-family overlap between a given community and either of the pathogens (fig. S9, A to D). However, by taking only the communities that contain *E. coli* to control for this effect, we also saw a strong correlation between a community's protein-family overlap with the pathogen and its colonization resistance in our in vitro assays (Fig. 4, A and B, and fig. S9). In other words, if the symbiont species or community encodes many of the same (or similar) proteins as the pathogen, it provides better colonization resistance. The same analysis for communities that lack *E. coli* is not

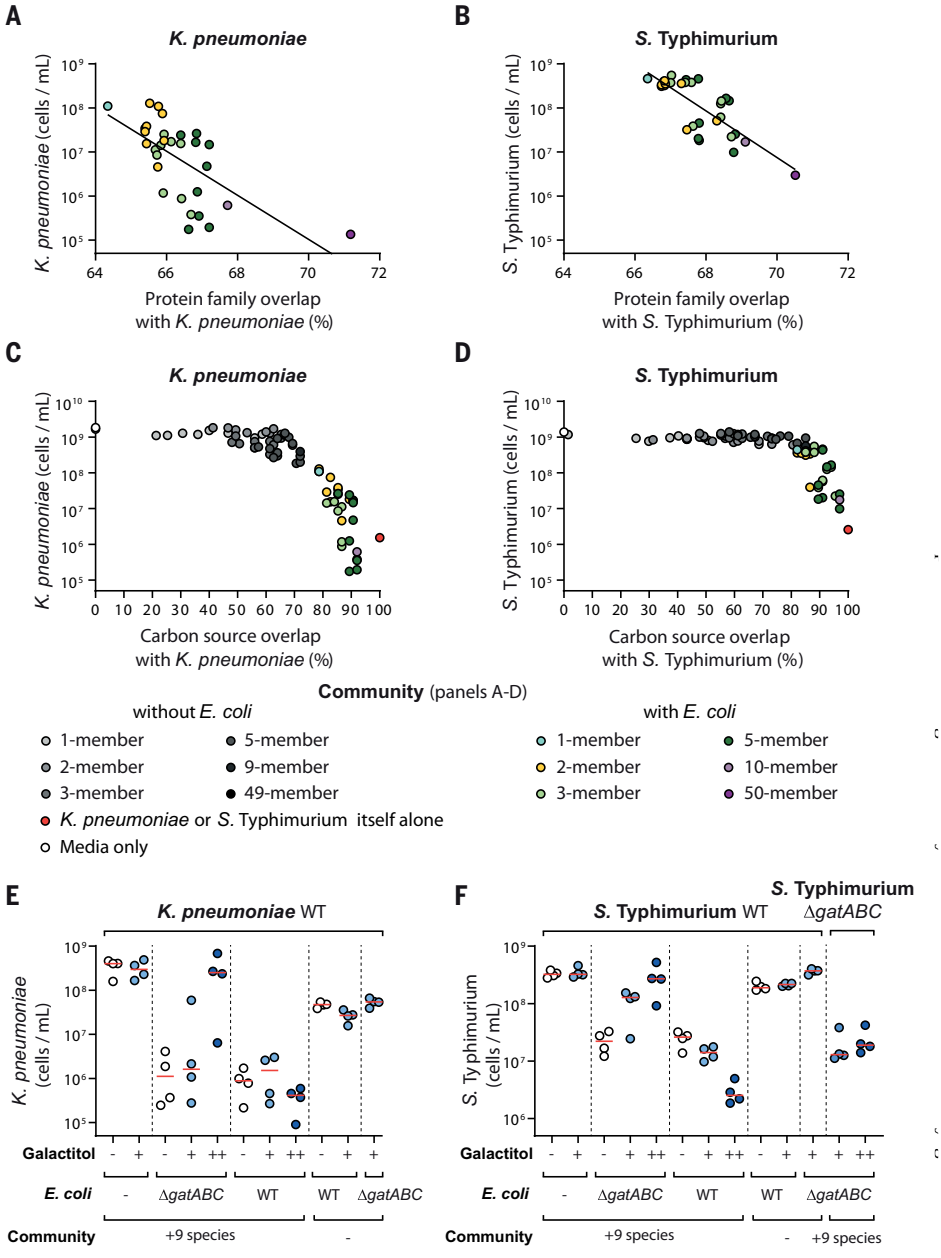


Fig. 4. Nutrient overlap can explain the role of ecological diversity and the effect of *E. coli* in colonization resistance. (**A** and **B**) Protein-family overlap is compared with the median pathogen abundance values for each community containing *E. coli* from Fig. 2, C and D. Diversity is visualized with a color gradient. Data for *K. pneumoniae* are shown in (A); data for *S.* Typhimurium are shown in (B). A line of best fit is shown from a linear regression on log-transformed data: R^2 = 0.4255 for *K. pneumoniae*; R^2 = 0.603 for *S.* Typhimurium (R^2, coefficient of determination); both slopes are significantly different from 0 according to an *F* test ($P < 0.0001$). Data for communities without *E. coli* are presented in fig. S9, E and F. (**C** and **D**) Overlap in carbon-source utilization plotted against the median pathogen abundance measurements from experimental communities in Fig. 2, C and D. Community carbon-source overlap is calculated by using an additive approach from carbon-source overlap of individual strains with measurement on Biolog AN Microplates (fig. S10). Diversity is visualized with a gradient of color (for *E. coli*–containing communities) or grayscale (for communities without *E. coli*). A control with the isogenic pathogen itself (100% overlap) is plotted in red. Data for *K. pneumoniae* are shown in (C); data for *S.* Typhimurium are shown in (D). (**E** and **F**) A private nutrient, galactitol, that could only be used by the WT *E. coli* strain and the pathogens but not by the other symbionts nor by an *E. coli* Δ*gatABC* mutant, was supplemented to the media, and the extended competition assay was performed as before. In all treatments, pathogen abundance was measured by flow cytometry after 48 hours of growth postpassage instead of the usual 24 hours. This change did not influence the control experiments without galactitol but proved informative because we found that the growth impacts of galactitol were relatively slow. Results for *K. pneumoniae* are shown in (E); results for *S.* Typhimurium are shown in (F). n = 3 to 4 biological replicates from independent experiments per treatment. Horizontal red lines show the median of the replicates. Light-blue circles show results with 0.1% galactitol supplementation ("+"); dark-blue circles show results with 1% galactitol supplementation ("++"). White circles (control) show results with no nutrient supplementation ("–"). "+9 species" indicates the 9 additional species in the 10 best-performing species for each respective pathogen ("–" indicates when *E. coli* is added alone). In (F), a Δ*gatABC* mutant of *S.* Typhimurium was used in addition to the WT pathogen to verify the dependency of colonization on a private nutrient.

informative because colonization resistance is consistently so low across all communities (Fig. 2).

Our genomic analyses suggest that communities that overlap highly with the pathogens in encoded functions provide the best colonization resistance. These analyses support our hypothesis that niche overlap is important for the ecological patterns that we observed in colonization resistance. One of the key drivers of niche overlap is resource competition (*39*, *40*), which is a known contributor to colonization resistance to *K. pneumoniae* and *S.* Typhimurium (*12*, *13*). We thus explored the role of nutrient competition by generating metabolic profiles for the two sets of 10 key symbiont species identified in the original screen against each pathogen (Fig. 1); for this we used Biolog AN MicroPlates, which profile the metabolic activity of each strain on 95 carbon sources (fig. S10). To cover the two sets of 10 species, we actually profiled only 16 strains because there were some parallels between the two sets of top-ranked species in the luminescence screen. We first established that there was a strong positive association between the protein family (genomic) and metabolic (Biolog) overlap of communities with the pathogens (fig. S11). We then assessed the ability of metabolic overlap to predict colonization resistance (Fig. 4, C and D). Colonization resistance was only observed once communities shared sufficiently high overlap in their carbon-source utilization profile with a pathogen. Moreover, communities with the greatest metabolic overlap with a pathogen provided the greatest colonization resistance. An important observation from these data is that it is not diversity per se that predicts colonization resistance; it is the overlap between the pathogen and the communities. This pattern is made clear by the observation that communities having different diversities but the same overlap appear

proximally in the plots (Fig. 4, C and D, neighboring points of different color).

Our data point to the importance of nutrient competition, and specifically of nutrient-utilization overlap between a community and a pathogen, as an explanation for the patterns that we observed in colonization resistance. To further support this conclusion, we performed experiments in which the pathogens were grown in cell-free (spent) media collected from different communities, which excluded cell-cell contact mechanisms as explanations for colonization resistance. Growing the pathogen in the spent media of *E. coli* and the 10-species communities recapitulated the patterns seen in the competition experiments, which is consistent with the effect of nutrient competition (fig. S12). As a final test, we sought a nutrient that can be used by the pathogens only and used it to perform nutrient supplementation experiments (Fig. 4, E and F). We identified galactitol from the Biolog plates (fig. S10). The pathogens can use this sugar alcohol, but it has the desirable property that it cannot be used by any of the symbionts in our focal 10-species communities except for *E. coli*. We engineered a strain of *E. coli* that lacks the transporter for cell import (*E. coli* Δ*gatABC* deletion mutant). By adding in galactitol to our standard media, we found that colonization resistance in a diverse community is lost if the pathogens can use the nutrient but *E. coli* cannot (Fig. 4, E and F). However, colonization resistance is restored when *E. coli* can use the nutrient. Moreover, if a pathogen is engineered so that it cannot use galactitol (*S.* Typhimurium Δ*gatABC* deletion mutant), colonization resistance is restored. These outcomes are exactly as expected if nutrient competition is the cause of colonization resistance.

Our data show that the ability of a microbiota community to consume nutrients required by a pathogen for growth underlies the colonization resistance that we observed. The nutrient blocking effect is a property of the entire community rather than of any one species alone. That colonization resistance is a community-level trait explains the importance of the ecological diversity and complexity (*22*) that we observed in our experiments. Despite considerable genomic and metabolic overlap with the pathogens, a species such as *E. coli* does not consume enough of the different nutrients available to the pathogens to block colonization. It is only in combination with other species that *E. coli* becomes effective at limiting pathogen growth.

Nutrient blocking can identify protective communities

Our experiments indicate that colonization resistance is an ecologically complex trait but that this complexity can be understood and predicted through a simple underlying principle. As an additional test of these findings, we used the nutrient blocking principle to predict community compositions that provide colonization resistance to a bacterial strain that was not present in our initial experiments. For this test, we chose an antimicrobial resistant (AMR) clinical *E. coli* strain that was isolated from the urine of a patient. AMR *E. coli* strains are a major current target for alternatives to antibiotics because members of this species have recently been found to be responsible for the most AMR-associated deaths of any bacterial species (*41*).

We first analyzed the AMR *E. coli* isolate on Biolog AN MicroPlates to assess its carbon-source utilization and compared this with the top-ranked strains from our initial luminescence screen (Fig. 1). We reasoned that these strains were a good place to start because *E. coli* is also a member of the *Enterobacteriaceae*, as are the two pathogens that were used to select the top-ranked strains. As expected, the AMR *E. coli* had the greatest protein overlap with the symbiont *E. coli* in our 16 strains, but additional strains were predicted to be required to restrict nutrient availability on the basis of the overlap needed to suppress the two pathogens (fig. S13). We next used the Biolog data to computationally assemble all possible communities of one, two, three, and five species from the 16 strains and calculated their resource-utilization overlap with the AMR *E. coli* (Fig. 5A). Again, in line with our findings, diversity improved the median resource-utilization overlap, but this depended strongly on the presence of the symbiont *E. coli*.

The simplest test of the importance of nutrient blocking is to remove the symbiont *E. coli* from a community and test the impact. Doing this for the community of all 16 strains confirmed the importance of *E. coli* for colonization resistance (Fig. 5B). However, we also tested our ideas on communities that contain *E. coli*. In these experiments, we identified communities predicted to have the highest and lowest overlap with the target strain at each diversity level. If there were ties in rank, communities were selected at random from those with the same level of overlap. We then used our extended competition assay (Fig. 1D) to test the ability of the AMR *E. coli* to invade the communities. As predicted by nutrient blocking, this assay revealed that increasing diversity leads to increased colonization resistance; critically, for each diversity level, the community predicted to resist the AMR *E. coli* consistently performed better in colonization resistance than the community predicted to do poorly (Fig. 5B). This result was clearest for the two- and three-species communities. For the five-species community, the best-performing community was only marginally better than the worst. We reasoned that this was because in these experiments, we were limited to choosing from 16 strains that were preselected for being relatively good competitors to *Enterobacteriaceae* (Fig. 1, B and C, and fig. S1).

To test the nutrient blocking principle more robustly, we selected from a wider range of possible species, using our set of 50 that we used in our in vitro and in vivo experiments (Figs. 2 and 3). Most of these species had not been characterized for their functioning in community-level colonization resistance other than in the 50-species treatment. We also used this set of experiments to test the power of the nutrient blocking principle to predict colonization resistance on the basis of genomic data alone. Rather than using the Biolog phenotypic assay, we thus returned to our measure of protein-family overlap, which calculated the overlap in all protein types between an invading strain and different communities. To do this, we only had to sequence the AMR *E. coli* clinical isolate because all other strains were sequenced. Using the same approach as we had for the Biolog predictions, we then assembled communities in silico that all contained the symbiont *E. coli* strain and in each case calculated their protein family overlap with the AMR *E. coli* (Fig. 5, C and D). As before, we chose communities with the lowest and highest overlap to the AMR *E. coli* across a range of diversities (randomly choosing communities if there were ties in rank) and experimentally assessed colonization resistance using the extended competition assay.

We again saw the importance of community diversity in these experiments. Moreover, despite using only genomic information and a much larger set of possible communities, we observed improvement in colonization resistance from the worst to best communities at each diversity level (Fig. 5E). Lastly, we evaluated our ability to select highly and poorly performing communities by assessing colonization resistance in additional five-species communities. At the five-species level, more than 200,000 communities with *E. coli* can be assembled from 50 species. We used our algorithm to sample approximately 50,000, and from these we identified four additional community compositions predicted to perform well and four predicted to perform poorly (to provide five of each class). In line with our predictions, the communities predicted to be colonization resistant showed a median 100-fold reduction in the abundance of the AMR *E. coli* compared with those predicted to be permissive (Fig. 5F).

Discussion

A key benefit of the microbiome is its ability to reduce the probability of infection through colonization resistance (*1*, *2*, *10*, *42*). In this study, we have used an ecological approach to understand the principles of colonization resistance in the gut microbiome. By screening a collection of human gut symbionts, we found that individual species were unable to provide

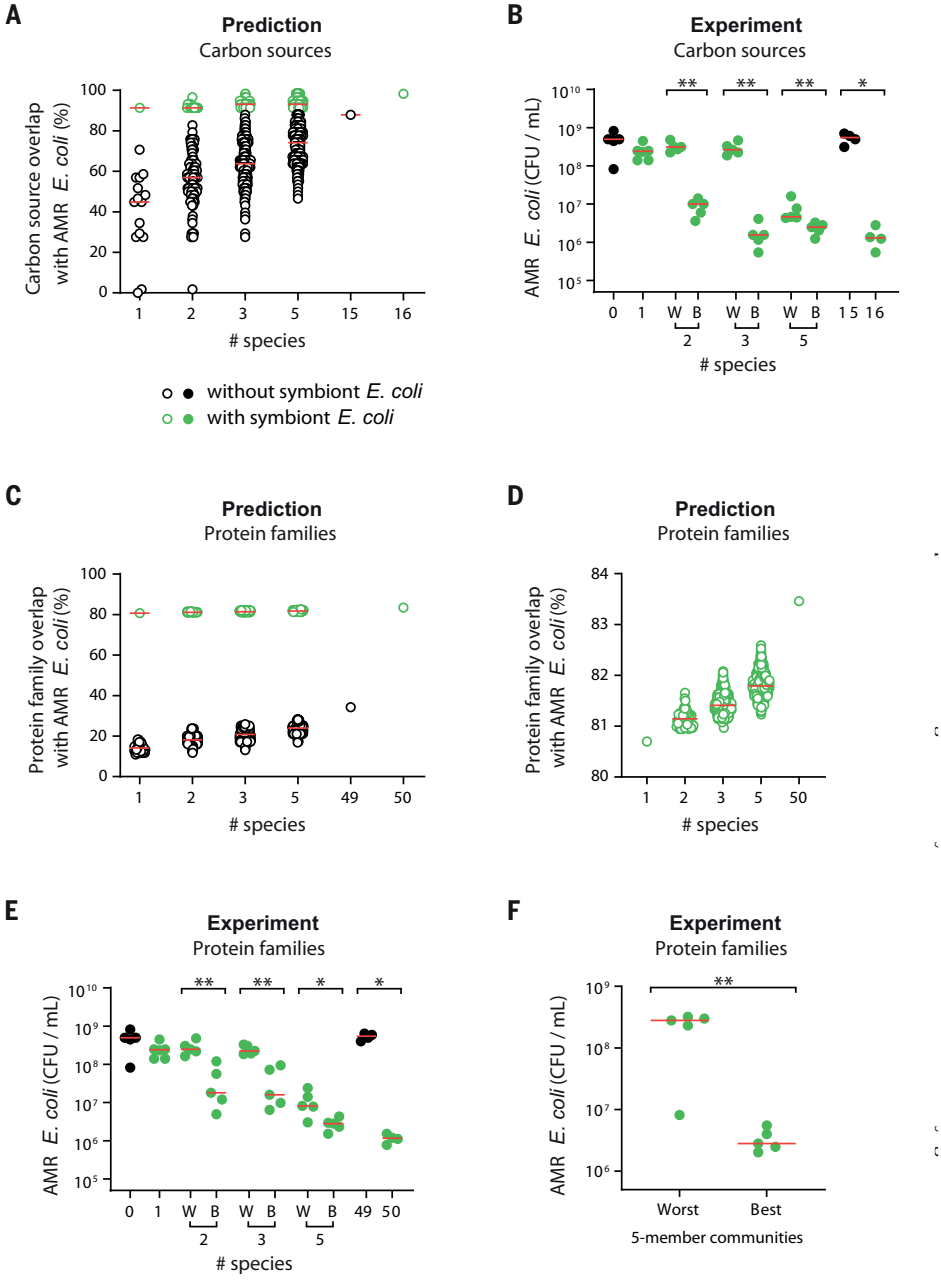


Fig. 5. Nutrient blocking predicts community colonization resistance. (**A**) In silico prediction of carbon-source overlap with the AMR *E. coli* strain for all possible combinations of symbiont communities at the indicated diversity levels. Each circle represents a different community. Communities containing the symbiont *E. coli* IAI1 are shown in green; communities without *E. coli* IAI1 are shown in black (predictions are shown as hollow circles; experimental data are shown as solid circles). Predicted carbon-source overlap is calculated by using an additive approach from carbon-source use of individual strains measured with Biolog AN MicroPlates (figs. S10 and S13). (**B**) Experimental test of in silico predictions in (A). The two *E. coli* IAI1–containing communities predicted to have the best (B) and worst (W) carbon-source overlap were picked at each diversity level and competed against AMR *E. coli* in the extended competition assay. A two-tailed Mann-Whitney *U* test was performed on community pairs ($P > 0.05$ = ns; $^{*}P < 0.05$; $P <$ **0.01) at the two-, three-, and five-species diversity levels. Red horizontal bars depict the median of each community tested. $n = 4$ to 5 biological replicates from independent experiments for each community. (**C** and **D**) In silico prediction of protein-family overlap with the AMR *E. coli* strain for a random subset ($n = 59{,}043$) of all possible symbiont communities at the two-, three-, and five-species diversity levels, as well as all individual species and the 49- and 50-species communities. Each circle represents a different community. Communities are selected from the species comprising the 50-member community. Communities containing *E. coli* IAI1 are shown in green; communities without *E. coli* IAI1 are shown in black. In (D), only the *E. coli*–containing communities are plotted to better visualize variation in protein-family overlap. (**E**) Experimental test of in silico predictions based on protein-cluster overlap in (C) and (D). The two *E. coli* IAI1–containing communities predicted to have the best (B) and worst (W) protein family overlap were picked at each diversity level (randomly selected for cases where there were multiple communities with the same overlap) and competed against AMR *E. coli* in the extended competition assay. Red horizontal bars depict the median of each community tested. $n = 5$ biological replicates from independent experiments for each community. A two-tailed Mann-Whitney *U* test was performed on community pairs ($P > 0.05$ = ns; $^{*}P < 0.05$; $^{**}P < 0.01$) at the two-, three-, and five-species diversity levels. (**F**) Experimental test of the predicted five-best and five-worst communities at the five-species diversity level, based on protein-family overlap with AMR *E. coli*. Each symbol represents the median of $n = 5$ biological replicates from independent experiments per community. Red horizontal bars depict the median of the predicted best and worst communities. A two-tailed Mann-Whitney *U* test was performed ($^{**}P < 0.01$). Community identities for (B), (E), and (F) are shown in table S7.

effective resistance to pathogens (Fig. 1) but that colonization resistance increased monotonically with ecological diversity (Figs. 2 and 3). Therefore, our work supports the general hypothesis that a more diverse microbiome can carry health benefits (*28*, *43*–*45*). Although much discussed, evidence for this hypothesis is typically based on correlations between microbiome diversity and health outcomes (*28*, *45*, *46*). Here, we provide experimental evidence that microbiome diversity can provide health benefits through an increased ability to protect against pathogens. Moreover, we have explained this pattern in terms of the importance of the overlap between the nutrient requirements of an invading pathogen and the resident community (Fig. 4).

We found that certain combinations of species display much greater colonization resistance together than when alone. These nonadditive effects mean that colonization resistance is formally a complex ecological trait in the canon of ecology (*22*). Such effects are often assumed to imply a complex network of interactions between species in which, for example, one symbiont species affects a second symbiont species and changes the way this second species interacts with a pathogen. However, consideration of nutrient competition, and particularly of the level of overlap between a pathogen and community, suggested that much simpler processes explain the complexity that we see. One species alone is not sufficient to strongly impact

pathogen growth, but rather a combination of species is required to limit nutrient access. We found that the combinations of species that make colonization-resistant communities can be very phylogenetically diverse. Generating resistant communities does not rest upon simply finding species closely related to a given pathogen—in our case, against *Enterobacteriaceae* members. Instead, a mixture of gram-positive and gram-negative species is often what performs best (Fig. 5F and table S7). Although both of our pathogens are members of the same family, they have different life histories (*19*–*21*). *S.* Typhimurium is a specialist gut pathogen, whereas *K. pneumoniae* is an opportunistic pathogen that typically causes no pathology in the gut itself, instead causing infections in other parts of the body. Consistent with this difference, we find that it is consistently easier to generate colonization resistance to *K. pneumoniae* than to *S.* Typhimurium. Nevertheless, the importance of ecological diversity and complexity is observed for both pathogens. We anticipate that the importance of ecological diversity and the principle of nutrient blocking will apply generally, given the widespread evidence that nutrient competition is important for diverse species in the microbiome, including many other pathogens (*9*, *47*–*55*).

By assembling a wide range of communities of defined compositions, we have been able to establish links between ecological diversity, complexity, and nutrient competition in colonization resistance. However, a limitation of this approach is that we have focused on communities of relatively low diversity. A question for the future is whether our findings will hold for the higher levels of diversity that can naturally occur in the human microbiome. Consistent with our findings, colonization risk with species such as *K. pneumoniae* is increased after antibiotic treatment that can lower species diversity in the microbiome (*56*). Moreover, *K. pneumoniae* carriage has been found to correlate negatively with the presence of *E. coli* in observational data from a neonate cohort, which may indicate competition between the two species (*57*). Nutrient competition is central to the patterns that we have described here. Colonization resistance can develop through additional mechanisms, which include toxin-mediated bacterial competition and effects via the host immune system (*1*, *10*). Our work does not exclude the potential importance of these additional mechanisms, such as direct killing of invading strains by members of the community, which can act in parallel to nutrient competition (*12*, *13*, *15*, *39*). Moreover, our predictions of resistant and permissive communities are not perfect (e.g., high-overlap communities in Fig. 4C and the outlier in Fig. 5F). However, it is notable that these deviations are in the direction of a community being more resistant than expected. This finding suggests that predictions based on nutrient blocking may often be conservative, with errors resulting in communities performing better than expected whenever other mechanisms of colonization resistance are at play.

In our experiments, colonization resistance is not the property of single microbiome species but instead the collective property of multiple species. Specifically, we found that the effect of a given symbiont species on a pathogen can be strongly dependent on whether other symbiont species are also present to consume nutrients that the pathogen needs. This finding suggests that one can use the idea of nutrient blocking to identify sets of microbiome species that will limit the growth of a target strain. As a proof of principle, we tested this idea for an AMR *E. coli* strain, which revealed that species sets can be successfully identified that collectively suppress an incoming strain (Fig. 5). We found that this can be done without specific information on available nutrients or the metabolism of the species under study. Instead, a measure of genomic overlap can be used as a proxy for niche overlap to assemble communities that perform nutrient blocking (Fig. 5, C to F). The human microbiome is dauntingly complex and has great potential for context-dependent effects. However, we found that microbiome complexity can arise through simple underlying principles, which gives promise to the goal of rationally designing microbiomes for better health.

Materials and methods

Bacterial strains and plasmids

A full list of bacterial strains used in this study is provided in table S1. The plasmids used in this study are listed in table S2. *Klebsiella pneumoniae* subsp. *pneumoniae* purchased from DSMZ (stock number 30104) was used in all experiments containing *K. pneumoniae*. *Salmonella enterica* serovar Typhimurium [strain SL1344 (*58*)] was used in all experiments containing *S.* Typhimurium. The 100 strains were chosen for being abundant and/or important members of the human gut microbiota that cover the key phylogenetic groups found in the human gut microbiota. In addition, all are type strains, have had their genome sequenced (which we rely on for the protein family analysis), and are known to be culturable in nutrient-rich, anaerobic medium (mGAM) (*59*, *60*), so were conducive to experimental work. Strains were kindly provided by Nassos Typas (EMBL Heidelberg, Germany), or were ordered from the German Collection of Microorganisms (DSMZ) or the American Type Culture Collection (ATCC; *Staphylococcus epidermidis*). The *Lactobacillus plantarum* strain was provided by the Department of Food and Nutritional Sciences, University of Reading. *K. pneumoniae* ATCC 700721 was provided by the Modernising Medical Microbiology research group, Nuffield Department of Medicine, University of Oxford. *E. coli* strains Z1331 and Z1269 were isolated from the feces of two healthy human donors in Switzerland (*33*). The ampicillin-resistant AMR *E. coli* strain is a urine clinical isolate provided by the Pathogen Bank at Nottingham University Hospitals NHS Trust.

Bacterial growth conditions

For engineering of *E. coli*, *K. pneumoniae*, or *S.* Typhimurium, strains were grown aerobically in Lysogeny broth (LB; Fisher Scientific) with the appropriate antibiotics (table S1) at 37°C, shaking at 220 rpm. All symbionts were cultured under anaerobic conditions (5% H_2, 5% CO_2, 90% N_2, <20 ppm O_2) in modified Gifu Anaerobic Medium (mGAM; Nissui Pharmaceuticals) broth buffered to pH 6.2 with 100-mM 2-(N-morpholino)ethanesulfonic acid (MES; Sigma-Aldrich). The redox indicator dye Resazurin (100 μg/liter media; Sigma-Aldrich) was added as a quality control check for the presence of oxygen (turns red when conditions are not sufficiently anaerobic). To prepare glycerol stocks, individual strains were first streaked onto mGAM agar (Nissui Pharmaceuticals) and grown under anaerobic conditions (if possible, as not all strains can grow as single colonies on agar-based media). The identity of each strain was confirmed using 16S rRNA sequencing (Sanger sequencing; Source Biosciences) using the primers oOPC-953 and oOPC-954 for most species, or g-Bifid-F and g-Bifid-R for Bifidobacteria (table S3). Single colonies were then inoculated in mGAM broth and stored at −70°C in mGAM with a final concentration of 25% glycerol.

Genetic engineering of bacterial strains

Luminescence and fluorescence plasmids were transformed into *K. pneumoniae* and *S.* Typhimurium using electroporation. Briefly, 5 ml overnight culture was washed three times with cold Milli-Q water before being concentrated in 500 μl cold Milli-Q water. 2 μl of plasmid was mixed with 100 μl of concentrated cells and electroporated (1.8 kV; 0.1-cm-gap cuvettes) before recovery in 1 ml prewarmed LB (1 hour at 37°C, shaking at 220 rpm) and plating on LB with the appropriate antibiotics.

Gene deletions were generated as described in (*61*). Briefly, 700 base pairs upstream and downstream of the region to be deleted were polymerase chain reaction (PCR) amplified (Phusion®, NEB) and inserted into the suicide vector pFOK, which was linearized using the restriction enzymes BamHI and EcoRI, using the NEBuilder® HiFi DNA Assembly (NEB). The plasmid was introduced into a diaminopimelic acid auxotroph *E. coli* strain (JKe201). After 6 hours of mating between the plasmid-containing donor *E. coli* strain and the recipient *E. coli*, *S.* Typhimurium, or *K. pneumoniae* strain, transconjugants were selected on LB

plates containing 50 μg/ml kanamycin. Counterselection was performed on no-salt LB plates supplemented with 0.5 μg/ml of anhydrous tetracycline and 20% of sucrose at 30°C. Mutants were screened by colony PCR and the sequence was verified using Sanger sequencing (Eurofins). Primers used for genetic engineering in this study are listed in table S3.

Luminescence screen

K. pneumoniae (DSM 30104) and *S.* Typhimurium (SL1344) carrying a low–copy number plasmid with P_{nptII} promoter-driven expression of *lux*CDBAE-*frp* [pRSJ-p_{nptII}::ilux; also known as the improved lux operon plasmid (*62*)] were used for the luminescence screen. Symbiont strains were tested in pairwise coculture with the pathogens. All strains tested were first grown in monoculture anaerobically in static mGAM broth buffered to pH 6.2 with 100 mM MES at 37°C. *K. pneumoniae* wild type (WT) and *S.* Typhimurium WT carrying the luminescence plasmid pRSJ-p_{nptII}::ilux were separately incubated in the same media overnight.

Competition assay

Once all strains had reached stationary phase (12 to 72 hours of growth depending on the species), 20 μl of each symbiont strain was dispensed in technical triplicates in a 96-well plate containing 160 μl of mGAM broth. 20 μL of *K. pneumoniae* pRSJ-p_{nptII}::ilux overnight culture or *S.* Typhimurium pRSJ-p_{nptII}::ilux overnight culture was then added to each well.

Ecological invasion assay

180 μl of each symbiont strain was dispensed in technical triplicates in a 96-well plate, and 20 μl *K. pneumoniae* pRSJ-p_{nptII}::ilux overnight culture or *S.* Typhimurium pRSJ-p_{nptII}::ilux overnight culture was added to each well.

96-well plates from both assays were incubated anaerobically for 6 hours at 37°C. Before luminescence measurement, 96-well plates were brought out of the anaerobic chamber and exposed to O_2 for 10 min because, critically, O_2 is needed for light production catalyzed by the enzyme luciferase. Luminescence at 515 to 575 nm and optical density at 600 nm (OD_{600}) were measured using a CLARIOstar Plus spectrometer (BMG Labtech). Each data point (fig. S1) represents the median of at least three independent biological replicates (that is, using different overnight cultures of the same strain in different 96-well plates).

Selection of the 10 best- and worst-ranked symbiont species from the screen

For each of the ecological invasion and competition assays, symbiont strains were ordered from best- to worst-performing. Symbiont strains were given two rankings: one from the competition assay and one from the invasion assay. The two rankings were summed and ordered from lowest sum (best overall performing) to highest sum (worst overall performing). From these ranks, the 10 best- and worst-ranked species were chosen. In order to focus on the ability of nonpathogenic strains to limit colonization by pathogens, we decided to focus for the community experiments on strains that have a category 1 safety level. Many common gut bacteria cause occasional infections and are therefore categorized as opportunistic pathogens with a category 2 safety level, which can blur the line between being protective and causing disease. Removing these strains took us from 100 to 54 strains. From these, we selected 50 species as our core set, removing some strains that were duplicates of the same species (two of the 50 differ at the subspecies level rather than the species level, but we use "species" for simplicity in the main text).

Construction of phylogenomic tree

The phylogenomic tree of strains in the luminescence screen (n = 100; Fig. 1, B and C, and table S1) was inferred using default settings based on 67 single-copy core genes using anvio v7.1 (*63*) and plotted using iTOL v6.7.5 (*64*).

Extended competition assay

All culturing was performed in a shaking incubator at 225 rpm and 37°C, and under anaerobic conditions. Symbiont strains were grown to stationary phase (12 to 72 hours, depending on the species), in Hungate tubes containing 5 ml mGAM broth and 100 mM MES and buffered to pH 6.2. *K. pneumoniae* and *S.* Typhimurium carrying the fluorescence plasmid pBC11 (YPet) (*65*) were incubated overnight with the appropriate antibiotics (table S1).

Once in stationary phase, monocultures of symbionts were passaged (100 μl culture into a new 5-ml tube of media) and grown for ~17 hours. Communities were assembled under anaerobic conditions into new 5-ml Hungate tubes (see "community preparation" section) and grown for 24 hours. Once grown, communities were invaded with 100 μl of the pathogen (10^6 cells/ml final concentration) in each Hungate tube containing 5 ml culture. Samples were taken immediately after addition of the pathogen and prepared for measurement using flow cytometry (day 0). After 24 hours growth, the invaded communities were sampled for flow cytometry (day 1) and 100 μl was passaged into new tubes of 5 ml mGAM broth. After a further 24 hours, the end-point communities were sampled and measured a final time (day 2).

For nutrient supplementation experiments (galactitol), 2.5 ml mGAM buffered with 100 mM MES to pH 6.2 was prepared and autoclaved in Hungate tubes. A 2× stock solution of filter-sterilized sugar in Milli-Q water (e.g., 0.2% or 2% galactitol) was prepared and deoxygenated in the anaerobic chamber. 2.5 ml of the stock solution was added into the Hungate tube to generate a 5-ml total volume with the appropriate concentration of the sugar. The extended competition experiment was performed as described above, except samples were passaged again into tubes with the appropriate concentration of the sugar. Samples were left to grow for 48 hours instead of the standard 24 hours to allow utilization of low-priority sugars, such as galactitol.

For experiments with the AMR *E. coli* clinical isolate, the experiment was performed as described above, except samples were analyzed using selective plating instead of flow cytometry (LB + 100 μg/ml ampicillin to select for the AMR *E. coli* strain; MacConkey agar to enumerate total *E. coli* densities).

Community preparation

All culturing was performed in a shaking incubator at 225 rpm and 37°C and under anaerobic conditions. Constituent monocultures were grown to stationary phase, passaged into new media, and grown overnight for approximately ~17 hours, and then combined into communities. The OD_{600} of the single species was measured, and the cultures were streaked anaerobically on mGAM agar to check for contamination. Communities ranged from 2 to 50 species in size and were assembled under anaerobic conditions.

In vitro

For the communities containing 2 to 10 species, equal volumes (100 μl) of each overnight culture were combined in a Hungate tube containing 5 ml mGAM to form a community. For the 49- and 50-strain communities, 20 μl of each single species was added, equating to a total volume of 1 ml. The communities were made by adding 1 ml of each strain anaerobically to a 50-ml falcon tube, mixing, and removing 1 ml to add to a fresh Hungate tube of 5 ml mGAM broth. Communities were grown in the shaking incubator (37°C, 225 rpm) for 24 hours before being invaded with 10^6 cells/ml of the pathogen.

In vivo

Species were grown separately, then OD_{600} of each culture was measured. For communities containing 2 to 10 species, the volume of culture containing 10^9 cells was calculated for each individual species, based on a standard curve of OD_{600} and flow cytometry quantification. This was added to a large falcon tube. The community mix was centrifuged (10 min, room temperature, 14,000 rpm), and the resulting pellet resuspended in 2 ml anaerobic phosphate-buffered saline (PBS) and kept on ice. 200 μl were gavaged into each mouse, corresponding to 10^8 cells of each species. For the 49- or 50-species communities, we used fewer cells of each species to avoid gavaging a large total cell density into mice (as done for the in vitro

experiments). The volume of culture containing 2×10^8 cells was calculated for each constituent species, and they were combined. The mixture was centrifuged and the pellet resuspended in 2 ml PBS. Mice were gavaged with 200 μl of the inoculum. Therefore, the density of bacteria in the 200-μl gavage was equivalent for the 50-species community as for the 10-species community (10^9 total cells). For all inoculums, samples were taken for plating anaerobically on mGAM agar to confirm the cell density.

Flow cytometry

Hungate tubes containing culture to be sampled were depressurized under anaerobic conditions using needles. Samples were removed using needles and syringes. 10 μl of each sample was diluted in 90 μl PBS + chloramphenicol 200 μg/ml, then 10 μl was added to a 96-well flat-bottomed plate already containing 80 μl PBS + chloramphenicol 200 μg/ml. The 96-well plate was left to shake at room temperature on the bench for 15 min to allow oxygen-dependent folding of YPet, while preventing further growth of the sample with the bacteriostatic antibiotic chloramphenicol. Next, all bacterial cells were fixed, permeabilized, and fluorescently stained by adding 90 μl PBS containing 4% paraformaldehyde, 0.4% Triton X-100, and 1 μg/ml 4′,6-diamidino-2-phenylindole (DAPI) to each sample well and incubating the plate in the dark for 1 hour at room temperature on a microplate shaker. Prior to the measurement, 20 μl AccuCheck counting beads (ThermoFisher- PBC100) was added to every sample well. Two wells containing 200 μl PBS were left between sample wells, in order to prevent cross-contamination between samples. 96-well plates were run in a flow cytometer (Attune NxT Autosampler with Attune NxT Flow Cytometer, ThermoFisher) set to: acquisition volume 50 μl, total draw volume 80 μl, total sample volume 200 μl, at a speed of 100 μl/min. Relevant spectral parameters were recorded by the flow cytometer, equipped with 405-nm, 488-nm, and 561-nm lasers. The following channels were used: DAPI staining, excitation 405 nm, emission 415 to 465 nm ("blue"); and YPet fluorescence, excitation 488 nm, emission 505 to 515 nm and 525 to 555 nm ("green").

Flow cytometry data was processed using FlowJo v10.8.1 software. "Beads" were gated on a linear (FSC-H; SSC-H) axis. "Bacteria" were gated on a logarithmic (FSC-H; SSC-H) axis. DAPI was gated within the "Bacteria" gate, on a logarithmic (Ex405-417; Ex405-495) axis. Within the DAPI gate, YPet was gated on a logarithmic (Ex488-503; Ex488-555) axis. Count statistics for beads, DAPI, and YPet were exported to Microsoft Excel and pathogen density (cells/ml) was calculated.

Mouse husbandry and experiments

Mouse experiments were performed with 6-to-8-weeks-old germ-free wild-type C57BL/6J female mice (RRID IMSR_JAX:000664) that were bred and maintained in germ-free isolators at the Kennedy Institute, University of Oxford for more than 5 years. During experiments, mice were housed as pairs or trios in sterile, individually ventilated Sentry SPP cages (Allentown) with enrichment, irradiated food, and autoclaved water. Cages of mice were disinfected in TecCare Ultra Hydrogen peroxide, peracetic acid solution before mice were handled using individually sterilized gloves (single use for each cage) in a biosafety cabinet that was disinfected with the same acid solution in between each treatment group. Cages of mice were randomly assigned to a group, and cages were opened in the same order each time. Researchers were not blinded because of the unbiased quantification methods of colony counting and metagenomic sequencing. Samples sizes were determined using standard practices of seven to eight mice per group distributed over two to three cages, and the experiment was repeated three times to allow experimental independence and sufficient power for nonparametric testing. Mouse experiments were performed in accordance with the UK Animal Scientific Procedures Act (1986) under a UK Home Office license (PPL 9127884) assessed by the Medical Sciences Division Animal Welfare and Ethical Review Body (AWERB) at the University of Oxford. Due to strictly defined humane endpoints, one mouse needed to be terminated early. In this case, we excluded data from that mouse.

Germ-free mice were gavaged with a 200-μl inoculum of symbionts (see "community preparation" section) at day −14 of the experiment and again on day −12 of the experiment (Fig. 3A, see scheme). On day 0, overnight cultures of the pathogens (*K. pneumoniae* and *S.* Typhimurium) grown aerobically in LB containing the appropriate antibiotics were washed three times with PBS and diluted. 10^6 colony forming units (CFUs) of the pathogen were given to each mouse as a 100-μl gavage. Feces were sampled daily after gavage of the pathogen. Fecal pellets were weighed and homogenized with a 5-mm stainless-steel bead in 1 ml PBS (samples were shaken vigorously in 2-ml Eppendorf tubes). CFUs were enumerated using selective plating (MacConkey or LB agar for *E. coli*; MacConkey + 50 μg/ml streptomycin for *S.* Typhimurium; LB + 50 μg/ml carbenicillin for *K. pneumoniae*). *E. coli*, *K. pneumoniae*, and *S.* Typhimurium can be differentiated by color on MacConkey agar (*S.* Typhimurium cannot utilize lactose) or morphology (*K. pneumoniae* produces a capsule). Mice were euthanized 4 days after infection with the pathogen.

S. Typhimurium triggers inflammation to bloom in the gut (*19*). To allow us to focus on early colonization events prior to triggering inflammation (i.e., colonization resistance), we used an avirulent strain (SL1344 Δ*invG* Δ*ssaV*) to avoid triggering inflammation, disease, and mortality (*66*).

DNA isolation from feces and metagenomic sequencing

Aliquots of the feces were frozen immediately after collection (prior to homogenization in PBS) and stored at −80°C until use. Fecal samples were thawed and resuspended in nuclease-free water before being transferred to a lysing matrix B tube (MP Biomedicals). Three rounds of bead beating were performed at 6 m/s for 40 s. Samples were centrifuged at high speed, and DNA in the supernatant was precipitated by adding sodium acetate (1/10 volume) and ice-cold ethanol (96 to 100%; equal volume) and left at −20°C overnight. Samples were centrifuged at high speed and the pellet was washed twice with 70% ethanol before being dried and resuspended in nuclease-free water.

Samples were further purified using an AMPure cleanup protocol. The samples were mixed with AMpure XP beads (Beckman Coulter) and incubated at room temperature for 5 min before being washed twice with 70% ethanol using a magnet to avoid removing the DNA bound to the AMPure beads. After air drying, nuclease-free water was added to the beads to allow collection of the purified DNA in the supernatant.

CosmosID performed metagenomic sequencing on isolated DNA. DNA libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina) and IDT Unique Dual Indexes. Prepared libraries were sequenced on the Illumina NovaSeq 6000 platform 2×150 bp (3 million reads per sample). CosmosID performed bioinformatic analysis according to proprietary methods on the raw data to generate fine-grained taxonomic and relative abundance estimates (fig. S5).

Spent media assay

Species were anaerobically inoculated, grown, passaged, and assembled into communities as described above. Communities were grown anaerobically in the shaking incubator (225 rpm, 37°C) for 96 hours to maximize depletion of nutrients utilizable by the community. Supernatants were prepared by centrifuging the 96-hour cultures (10 min, 4400 rpm, room temperature) and filter-sterilizing the supernatant (0.2-μm filter). All preparation of supernatants were done under anaerobic conditions. Two of each of the communities were assembled so that for each community, one Hungate tube containing 5 ml supernatant and another Hungate tube containing 2.5 ml supernatant and 2.5 ml mGAM (the resupplemented treatment) were prepared. The supernatants were invaded with 10^6 cells/ml of fluorescent pathogen, sampled, and incubated for 24 hours as before. Samples were taken on days 0 and 1.

Nutrient utilization overlap using Biolog assays

Utilization of carbon sources by symbionts and the pathogens were assessed using Biolog AN Microplates (Biolog) according to the protocol of the manufacturer. Briefly, species were grown in mGAM and passaged (as detailed above). The cultures were centrifuged anaerobically and washed twice in anaerobic PBS. The samples were concentrated, and an aliquot was taken to measure the OD_{600}. The cells were diluted or concentrated such that a 200-μl aliquot of the cells would correspond to 65% transmittance (OD 0.187; the density outlined by the manufacturer) in the 14-ml volume of AN inoculating fluid provided. The concentrated cells were added to the inoculating fluid aerobically (a small amount of O_2 is needed to oxidize the buffer), and 100 μl was aliquoted into each well of the microplate. After 10 min in aerobic conditions, the plates were put in an airtight container with a GasPak EZ Container System Sachet to generate hydrogen-free anaerobic conditions. Samples were incubated at 35°C for 24 hours before measurement at 590 nm. Since *S.* Typhimurium SL1344 is a histidine auxotroph, a prototrophic strain was made using P22 transduction (*67*) of the allele from a prototrophic strain of *S.* Typhimurium (ATCC 14028S). Positive clones were selected for on M9 minimal media with 50 μg/ml streptomycin and restreaked three times on LB plates to ensure that phages were removed. Growth in the absence of exogenous histidine was validated by streaking on M9 media.

For each plate, the absorbance reading at 590 nm was subtracted from the blank (no carbon source control in well A1). Each strain generates a different background signal, so it is important to do this on each plate individually. Each species was measured as three independent biological replicates and the median absorbance value was taken for each carbon source. We used a thresholding approach to determine if a strain can metabolize a given carbon source (defined as Abs 590 nm >0.1 median, after blank subtraction; fig. S10).

To predict overlap with the pathogens, whether a given species can use a carbon source was compared with the pathogen. We assessed which percentage of carbon sources that the pathogen can use can also be used by the symbiont. To calculate overlap between pathogens and communities, we used an additive calculation approach, where if a species is contained in a community that can use a carbon source, the entire community can use the carbon source. For simplicity, we did not treat cases where multiple species use the same nutrients within the community differently than if a given nutrient is only covered by one species.

Genomic analysis using protein overlap

Genomic information from 50 symbiont species and two pathogens (*S.* Typhimurium, *K. pneumoniae*) were retrieved from the PATRIC database (*68*) (table S1). On this set, we applied a cluster-based analysis that groups proteins encoded in genomes into PATRIC global protein families (*69*). A majority of symbiont species proteins obtained a protein-family designation (clustering rate of 97.77%; 153,771 of 157,283) with protein families being populated by proteins from 2.44 genomes on average (fig. S7).

Based on this set, we drew protein family designations for focal symbiont communities and calculated pathogen/community overlap using the same approach as for the Biolog plates above. That is, for each protein family encoded in the pathogen's genome, we checked if it overlapped with any protein families encoded in the community.

Whole-genome sequencing

An overnight culture of *E. coli* 19Y000018 (the AMR clinical urine isolate) was prepared in LB containing 100 μg/ml ampicillin. A 1-ml pellet was taken, and DNA was extracted using the same ethanol precipitation and AMPure clean-up protocol detailed above in the section "DNA isolation from feces." Source BioScience performed whole-genome Illumina sequencing on a NovaSeq 6000 to generate 10 million 150 bp paired-end reads. Genome assembly was performed with Unicycler v0.4.8 using default settings and subjected to protein-family annotation as described in "Genomic analysis using protein overlap."

Predictions based on Biologs and protein families

Biolog-based predictions

The carbon-source utilization profile of the AMR *E. coli* clinical isolate was assessed using Biolog AN Microplates as detailed above. Next, all possible communities made of one, two, three, and five species were identified from the 16 symbiont strains that were previously analyzed for carbon-source utilization. For each community, the carbon-source utilization overlap was calculated (Fig. 5A; as detailed above). At each diversity level, communities were sorted by overlap to the AMR *E. coli* clinical isolate. Given that the symbiont *E. coli* is needed for colonization resistance in our experiments, we only included communities where the symbiont *E. coli* is present. At each diversity level, we chose the highest-ranked community (defined "predicted best") and lowest-ranked community (defined "predicted worst"). If there were ties in rank, one community was randomly chosen. These communities were experimentally assessed for colonization resistance using the extended competition in vitro assay as detailed above.

Protein family-based predictions

We used the whole-genome sequence of the AMR *E. coli* isolate to calculate protein-family overlap to communities of symbionts drawn from our 50-species community. As for the Biolog-based predictions, we restricted the analysis to communities that contained the symbiont *E. coli* strain. Specifically, we generated 100,000 different communities randomly, each containing the symbiont *E. coli* and a total of 2, 3, 5, or 10 species from the 50-species pool. We ranked the communities according to their pathogen overlap at each diversity level and selected the highest- and lowest-ranked communities for experimental validation. If ties occurred, we chose the communities at random, as for the Biolog-based predictions.

Statistical analysis

All graphs were made and statistical analysis carried out in Prism v9.4.1 (GraphPad). Protein-family overlap analyses and protein family–based predictions were computed in R version v4.0.5 (*70*). Figure legends indicate the statistical test used and the sample sizes. Nonparametric tests were used to avoid the assumptions of normality.

Ethical statement

Mouse experiments were performed in accordance with the UK Animal Scientific Procedures Act (1986) under a UK Home Office license (PPL 9127884) assessed by the Medical Sciences Division Animal Welfare and Ethical Review Body (AWERB) at the University of Oxford. Mice were euthanized using cervical dislocation followed by exsanguination as a confirmation of death.

REFERENCES AND NOTES

1. M. T. Sorbara, E. G. Pamer, Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them. *Mucosal Immunol.* **12**, 1–9 (2019). doi: 10.1038/s41385-018-0053-0; pmid: 29988120
2. A. Jacobson *et al.*, A Gut Commensal-Produced Metabolite Mediates Colonization Resistance to Salmonella Infection. *Cell Host Microbe* **24**, 296–307.e7 (2018). doi: 10.1016/j.chom.2018.07.002; pmid: 30057174
3. E. M. Velazquez *et al.*, Endogenous Enterobacteriaceae underlie variation in susceptibility to Salmonella infection. *Nat. Microbiol.* **4**, 1057–1064 (2019). doi: 10.1038/s41564-019-0407-8; pmid: 30911125
4. S. Y. Wotzka *et al.*, Escherichia coli limits Salmonella Typhimurium infections after diet shifts and fat-mediated microbiota perturbation in mice. *Nat. Microbiol.* **4**, 2164–2174 (2019). doi: 10.1038/s41564-019-0568-5; pmid: 31591555
5. R. P. Sequeira, J. A. K. McDonald, J. R. Marchesi, T. B. Clarke, Commensal Bacteroidetes protect against Klebsiella pneumoniae colonization and transmission through IL-36 signalling. *Nat. Microbiol.* **5**, 304–313 (2020). doi: 10.1038/s41564-019-0640-1; pmid: 31907407
6. C. G. Buffie *et al.*, Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517**, 205–208 (2015). doi: 10.1038/nature13828; pmid: 25337874
7. S. G. Kim *et al.*, Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. *Nature* **572**, 665–669 (2019). doi: 10.1038/s41586-019-1501-z; pmid: 31435014
8. M. X. Byndloss *et al.*, Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science* **357**, 570–575 (2017). doi: 10.1126/science.aam9949; pmid: 28798125
9. R. A. Oliveira *et al.*, Klebsiella michiganensis transmission enhances resistance to Enterobacteriaceae gut invasion by nutrition competition. *Nat. Microbiol.* **5**, 630–641 (2020). doi: 10.1038/s41564-019-0658-4; pmid: 31959968

10. G. Caballero-Flores, J. M. Pickard, G. Núñez, Microbiota-mediated colonization resistance: Mechanisms and regulation. *Nat. Rev. Microbiol.* **21**, 347–360 (2023). doi: 10.1038/s41579-022-00833-7; pmid: 36539611
11. E. T. Granato, T. A. Meiller-Legrand, K. R. Foster, The Evolution and Ecology of Bacterial Warfare. *Curr. Biol.* **29**, R521–R537 (2019). doi: 10.1016/j.cub.2019.04.024; pmid: 31163166
12. C. Eberl *et al*., E. coli enhance colonization resistance against Salmonella Typhimurium by competing for galactitol, a context-dependent limiting carbon source. *Cell Host Microbe* **29**, 1680–1692.e7 (2021). doi: 10.1016/j.chom.2021.09.004; pmid: 34610296
13. L. Osbelt *et al*., Klebsiella oxytoca causes colonization resistance against multidrug-resistant K. pneumoniae in the gut via cooperative carbohydrate competition. *Cell Host Microbe* **29**, 1663–1679.e7 (2021). doi: 10.1016/j.chom.2021.09.003; pmid: 34610293
14. S. Caballero *et al*., Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. *Cell Host Microbe* **21**, 592–602.e4 (2017). doi: 10.1016/j.chom.2017.04.002; pmid: 28494240
15. S. Brugiroux *et al*., Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. *Nat. Microbiol.* **2**, 16215 (2016). doi: 10.1038/nmicrobiol.2016.215; pmid: 27869789
16. A. G. Cheng *et al*., Design, construction, and in vivo augmentation of a complex gut microbiome. *Cell* **185**, 3617–3636.e19 (2022). doi: 10.1016/j.cell.2022.08.003; pmid: 36070752
17. S. Widder *et al*., Challenges in microbial ecology: Building predictive understanding of community function and dynamics. *ISME J.* **10**, 2557–2568 (2016). doi: 10.1038/ismej.2016.45; pmid: 27022995
18. E. Tacconelli *et al*., Discovery, research, and development of new antibiotics: The WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018). doi: 10.1016/S1473-3099(17)30753-3; pmid: 29276051
19. B. Stecher *et al*., Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. *PLOS Biol.* **5**, e244 (2007). doi: 10.1371/journal.pbio.0050244; pmid: 17760501
20. S. E. Majowicz *et al*., The global burden of nontyphoidal Salmonella gastroenteritis. *Clin. Infect. Dis.* **50**, 882–889 (2010). doi: 10.1086/650733; pmid: 20158401
21. C. L. Gorrie *et al*., Gastrointestinal Carriage Is a Major Reservoir of Klebsiella pneumoniae Infection in Intensive Care Patients. *Clin. Infect. Dis.* **65**, 208–215 (2017). doi: 10.1093/cid/cix270; pmid: 28369261
22. M. M. Mayfield, D. B. Stouffer, Higher-order interactions capture unexplained complexity in diverse communities. *Nat. Ecol. Evol.* **1**, 62 (2017). doi: 10.1038/s41559-016-0062; pmid: 28812740
23. Y. Litvak, A. J. Bäumler, The founder hypothesis: A basis for microbiota resistance, diversity in taxa carriage, and colonization resistance against pathogens. *PLOS Pathog.* **15**, e1007563 (2019). doi: 10.1371/journal.ppat.1007563; pmid: 30789972
24. K. Z. Coyte, S. Rakoff-Nahoum, Understanding Competition and Cooperation within the Mammalian Gut Microbiome. *Curr. Biol.* **29**, R538–R544 (2019). doi: 10.1016/j.cub.2019.04.017; pmid: 31163167
25. O. Manor *et al*., Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat. Commun.* **11**, 5206 (2020). doi: 10.1038/s41467-020-18871-1; pmid: 33060586
26. C. A. Lozupone, J. I. Stombaugh, J. I. Gordon, J. K. Jansson, R. Knight, Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012). doi: 10.1038/nature11550; pmid: 22972295
27. C. Tropini *et al*., Transient Osmotic Perturbation Causes Long-Term Alteration to the Gut Microbiota. *Cell* **173**, 1742–1754.e17 (2018). doi: 10.1016/j.cell.2018.05.008; pmid: 29906449
28. K. V. Johnson, P. W. Burnet, Microbiome: Should we diversify from diversity? *Gut Microbes* **7**, 455–458 (2016). doi: 10.1080/19490976.2016.1241933; pmid: 27723427
29. A. Baichman-Kass, T. Song, J. Friedman, Competitive interactions between culturable bacteria are highly non-additive. *eLife* **12**, e83398 (2023). doi: 10.7554/eLife.83398; pmid: 36852917
30. I. Billick, T. J. Case, Higher Order Interactions in Ecological Communities: What Are They and How Can They be Detected? *Ecology* **75**, 1529–1543 (1994). doi: 10.2307/1939614
31. A. L. Gould *et al*., Microbiome interactions shape host fitness. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11951–E11960 (2018). doi: 10.1073/pnas.1809349115; pmid: 30510004
32. W. B. Ludington, Higher-order microbiome interactions and how to find them. *Trends Microbiol.* **30**, 618–621 (2022). doi: 10.1016/j.tim.2022.03.011; pmid: 35469711
33. S. Y. Wotzka *et al*., Microbiota stability in healthy individuals after single-dose lactulose challenge-A randomized controlled study. *PLOS ONE* **13**, e0206214 (2018). doi: 10.1371/journal.pone.0206214; pmid: 30359438
34. S. B. Formal, G. J. Dammin, E. H. Labrec, H. Schneider, Experimental Shigella infections: Characteristics of a fatal infection produced in guinea pigs. *J. Bacteriol.* **75**, 604–610 (1958). doi: 10.1128/jb.75.5.604-610.1958; pmid: 13538931
35. F. R. Blattner *et al*., The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453–1462 (1997). doi: 10.1126/science.277.5331.1453; pmid: 9278503
36. M. Ackermann *et al*., Self-destructive cooperation mediated by phenotypic noise. *Nature* **454**, 987–990 (2008). doi: 10.1038/nature07067; pmid: 18719588
37. E. Gül *et al*., The microbiota conditions a gut milieu that selects for wild-type Salmonella Typhimurium virulence. *PLOS Biol.* **21**, e3002253 (2023). doi: 10.1371/journal.pbio.3002253; pmid: 37651408
38. R. J. Gibbons, S. S. Socransky, B. Kapsimalis, Establishment of Human Indigenous Bacteria in Germ-Free Mice. *J. Bacteriol.* **88**, 1316–1323 (1964). doi: 10.1128/jb.88.5.1316-1323.1964; pmid: 14234787
39. R. Freter, H. Brickner, M. Botney, D. Cleven, A. Aranki, Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. *Infect. Immun.* **39**, 676–685 (1983). doi: 10.1128/iai.39.2.676-685.1983; pmid: 6339388
40. A. Wagner, Competition for nutrients increases invasion resistance during assembly of microbial communities. *Mol. Ecol.* **31**, 4188–4203 (2022). doi: 10.1111/mec.16565; pmid: 35713370
41. C. J. L. Murray *et al*., Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *Lancet* **399**, 629–655 (2022). doi: 10.1016/S0140-6736(21)02724-0; pmid: 35065702
42. M. R. McLaren, B. J. Callahan, Pathogen resistance may be the principal evolutionary advantage provided by the microbiome. *Philos. Trans. R. Soc. London Ser. B* **375**, 20190592 (2020). doi: 10.1098/rstb.2019.0592; pmid: 32772671
43. M. J. Blaser, The theory of disappearing microbiota and the epidemics of chronic diseases. *Nat. Rev. Immunol.* **17**, 461–463 (2017). doi: 10.1038/nri.2017.77; pmid: 28749457
44. M. Fassarella *et al*., Gut microbiome stability and resilience: Elucidating the response to perturbations in order to modulate gut health. *Gut* **70**, 595–605 (2021). doi: 10.1136/gutjnl-2020-321747; pmid: 33051190
45. H. C. Wastyk *et al*., Gut-microbiota-targeted diets modulate human immune status. *Cell* **184**, 4137–4153.e14 (2021). doi: 10.1016/j.cell.2021.06.019; pmid: 34256014
46. E. Le Chatelier *et al*., Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
47. S. E. Winter *et al*., Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature* **467**, 426–429 (2010). doi: 10.1038/nature09415; pmid: 20864996
48. K. M. Pruss, J. L. Sonnenburg, C. difficile exploits a host metabolite produced during toxin-mediated disease. *Nature* **593**, 261–265 (2021). doi: 10.1038/s41586-021-03502-6; pmid: 33911281
49. L. Maier *et al*., Microbiota-derived hydrogen fuels Salmonella typhimurium invasion of the gut ecosystem. *Cell Host Microbe* **14**, 641–651 (2013). doi: 10.1016/j.chom.2013.11.002; pmid: 24331462
50. M. L. Jenior, J. L. Leslie, V. B. Young, P. D. Schloss, *Clostridium difficile* Colonizes Alternative Nutrient Niches during Infection across Distinct Murine Gut Microbiomes. *mSystems* **2**, e00063–17 (2017). doi: 10.1128/mSystems.00063-17; pmid: 28761936
51. K. M. Ng *et al*., Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–99 (2013). doi: 10.1038/nature12503; pmid: 23995682
52. A. W. Hudson, A. J. Barnes, A. S. Bray, D. A. Ornelles, M. A. Zafar, Klebsiella pneumoniae l-Fucose Metabolism Promotes Gastrointestinal Colonization and Modulates Its Virulence Determinants. *Infect. Immun.* **90**, e0020622 (2022). doi: 10.1128/iai.00206-22; pmid: 36129299
53. A. G. Jimenez, M. Ellermann, W. Abbott, V. Sperandio, Diet-derived galacturonic acid regulates virulence and intestinal colonization in enterohaemorrhagic Escherichia coli and Citrobacter rodentium. *Nat. Microbiol.* **5**, 368–378 (2020). doi: 10.1038/s41564-019-0641-0; pmid: 31873206
54. F. C. Pereira *et al*., Rational design of a microbial consortium of mucosal sugar utilizers reduces Clostridiodes difficile colonization. *Nat. Commun.* **11**, 5104 (2020). doi: 10.1038/s41467-020-18928-1; pmid: 33037214
55. E. Gül *et al*., Differences in carbon metabolic capacity fuel co-existence and plasmid transfer between Salmonella strains in the mouse gut. *Cell Host Microbe* **31**, 1140–1153.e3 (2023). doi: 10.1016/j.chom.2023.05.029; pmid: 37348498
56. N. Raffelsberger *et al*., Gastrointestinal carriage of Klebsiella pneumoniae in a general adult population: A cross-sectional study of risk factors and bacterial genomic diversity. *Gut Microbes* **13**, 1939599 (2021). doi: 10.1080/19490976.2021.1939599; pmid: 34182896
57. T. Mäklin *et al*., Strong pathogen competition in neonatal gut colonisation. *Nat Commun.* **13**, 7417 (2022). doi: 10.1038/s41467-022-35178-5
58. S. K. Hoiseth, B. A. Stocker, Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. *Nature* **291**, 238–239 (1981). doi: 10.1038/291238a0; pmid: 7015147
59. L. Maier *et al*., Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628 (2018). doi: 10.1038/nature25979; pmid: 29555994
60. M. Tramontano *et al*., Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nat. Microbiol.* **3**, 514–522 (2018). doi: 10.1038/s41564-018-0123-9; pmid: 29556107
61. F. R. Cianfanelli, O. Cunrath, D. Bumann, Efficient dual-negative selection for bacterial genome editing. *BMC Microbiol.* **20**, 129 (2020). doi: 10.1186/s12866-020-01819-2; pmid: 32448155
62. R. Soldan *et al*., From macro to micro: A combined bioluminescence-fluorescence approach to monitor bacterial localization. *Environ. Microbiol.* **23**, 2070–2085 (2021). doi: 10.1111/1462-2920.15296; pmid: 33103833
63. A. M. Eren *et al*., Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* **6**, 3–6 (2021). doi: 10.1038/s41564-020-00834-3; pmid: 33349678
64. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021). doi: 10.1093/nar/gkab301; pmid: 33885785
65. O. Cunrath, D. Bumann, Host resistance factor SLC11A1 restricts *Salmonella* growth through magnesium deprivation. *Science* **366**, 995–999 (2019). doi: 10.1126/science.aax7898; pmid: 31753999
66. S. Hapfelmeier *et al*., The Salmonella pathogenicity island (SPI)-2 and SPI-1 type III secretion systems allow Salmonella serovar typhimurium to trigger colitis via MyD88-dependent and MyD88-independent mechanisms. *J. Immunol.* **174**, 1675–1685 (2005). doi: 10.4049/jimmunol.174.3.1675; pmid: 15661931
67. N. L. Sternberg, R. Maurer, Bacteriophage-mediated generalized transduction in Escherichia coli and Salmonella typhimurium. *Methods Enzymol.* **204**, 18–43 (1991). doi: 10.1016/0076-6879(91)04004-8; pmid: 1943777
68. A. R. Wattam *et al*., Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* **45**, D535–D542 (2017). doi: 10.1093/nar/gkw1017; pmid: 27899627
69. J. J. Davis *et al*., PATtyFams: Protein Families for the Microbial Genomes in the PATRIC Database. *Front. Microbiol.* **7**, 118 (2016). doi: 10.3389/fmicb.2016.00118; pmid: 26903996
70. R Development Core Team: R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2021). https://www.R-project.org/.
71. F. Spragge *et al*., Microbiome diversity protects against pathogens by nutrient blocking, Dryad (2023); https://doi.org/10.5061/dryad.pnvx0k6v8.

ACKNOWLEDGMENTS

We are indebted to members of the Foster lab for discussion and to E. Slack, K. Coyte, A. Weiss, and W.-D. Hardt for feedback on the manuscript. We thank F. Powrie and the Oxford Centre for Microbiome Studies for germ-free mice and gnotobiotic mouse

work. The Heidelberg strains were a gift from Nassos Typas, EMBL Heidelberg. The AMR *E. coli* strain 19Y000018 was from Nottingham University Hospitals Pathogen Bank (https://www.nuh.nhs.uk/pathogen-industry/). **Funding:** F.S. was supported by a BBSRC Studentship. E.B. was supported by SNSF postdoc mobility fellowships (P2EZP3_199916 and P500PB_210941). M.T.J. was supported by the Human Frontier Science Program (LT000798/2020). This work was supported by Wellcome Trust Investigator award 209397/Z/17/Z and by European Research Council Grant 787932 to K.R.F. **Author contributions:** Conceptualization: K.R.F. Methodology: F.S., E.B., M.T.J., E.B.N.A., C.F.P., X.W., L.P., and O.C. Investigation: F.S., E.B., M.T.J., and O.C. Visualization: F.S., E.B., and M.T.J. Funding acquisition: K.R.F. Supervision: E.B., O.C., and K.R.F. Writing – original draft: F.S., E.B., and K.R.F. Writing – review and editing: F.S., E.B., M.T.J., C.F.P., X.W., L.P., O.C., and K.R.F. **Competing interests:** K.R.F. holds equity in Postbiotics Plus Research. **Data and materials availability:** All data used to generate the plots are available at Dryad (*71*). All code is available at GitHub (https://github.com/MartinTJahn/Nutrient_blocking). Metagenomic sequencing and whole-genome sequencing of *E. coli* 0018 (NCBI accession no. JAVXZX000000000) has been deposited in SRA under BioProject PRJNA1021490. There is no restriction on use of the data, materials, or code, with the exception of *E. coli* strain 19Y000018, which is protected by an MTA and requires permission from the Nottingham University Hospitals Pathogen Bank. A Materials Design Analysis Reporting checklist is supplied with the publication. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by Wellcome (Investigator Award 209397/Z/17/Z), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj3502
Figs. S1 to S13
Tables S1 to S7
References (*72*–*74*)
MDAR Reproducibility Checklist

Submitted 22 June 2023; accepted 1 November 2023
10.1126/science.adj3502

RESEARCH ARTICLE

MESOSCOPIC PHYSICS

Emission and coherent control of Levitons in graphene

A. Assouline[1]†, L. Pugliese[1,2]†, H. Chakraborti[1]†, Seunghun Lee[3], L. Bernabeu[1,2], M. Jo[1], K. Watanabe[4], T. Taniguchi[4], D. C. Glattli[1], N. Kumada[5], H.-S. Sim[3], F. D. Parmentier[1], P. Roulleau[1]*

Flying qubits encode quantum information in propagating modes instead of stationary discrete states. Although photonic flying qubits are available, the weak interaction between photons limits the efficiency of conditional quantum gates. Conversely, electronic flying qubits can use Coulomb interactions, but the weaker quantum coherence in conventional semiconductors has hindered their realization. In this work, we engineered on-demand injection of a single electronic flying qubit state and its manipulation over the Bloch sphere. The flying qubit is a Leviton propagating in quantum Hall edge channels of a high-mobility graphene monolayer. Although single-shot qubit readout and two-qubit operations are still needed for a viable manipulation of flying qubits, the coherent manipulation of an itinerant electronic state at the single-electron level presents a highly promising alternative to conventional qubits.

Flying qubit experiments rely on the ability to encode information into a propagating state of a single photon or electron excitation, manipulate the information, and read it after operations (*1*). Photon flying qubits evolve in the vacuum, which substantially reduces decoherence processes. Conversely, electron flying qubits naturally experience a strong and tunable Coulomb interaction, which leads to easier two-qubit operations but gives rise to finite decoherence. Electronic flying qubits can benefit from recent breakthroughs of electron quantum optics in GaAs heterostructures, including the Mach-Zehnder interferometer (MZI) (*2*–*4*), Hong-Ou-Mandel experiments (*5*, *6*), robust high-fidelity single-electron sources based on Levitons (voltage pulses with a Lorentzian profile enabling pure-electron excitation without any hole) (*6*, *7*), and the demonstration of single-electron quantum tomography (*8*). Nevertheless, a basic quantum manipulation of an on-demand and propagating single electron is still missing (*9*), primarily owing to the short coherence length (*3*) of excited electrons in conventional semiconductors. This problem could be solved by using graphene, a two-dimensional atomically thin material, which shows outstanding coherence properties under relatively large bias (*10*, *11*). Owing to its valley pseudospin degrees of freedom, graphene provides a very promising platform.

[1]SPEC, CEA, CNRS, Université Paris-Saclay, CEA Saclay, 91191 Gif sur Yvette Cedex, France. [2]Université Paris-Saclay, CNRS, Centrale Supélec, 91191 Gif sur Yvette Cedex, France. [3]Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea. [4]National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [5]NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi 243-0198, Japan.
*Corresponding author. Email: preden.roulleau@cea.fr
†These authors contributed equally to this work.

Recently, it was shown that the valley degrees of freedom in graphene can be addressed electrostatically (*12*–*17*). In particular, coherent and tunable electronic beam splitters, which couple quantum Hall edge channels with opposite valley polarizations, were formed (*10*, *18*, *19*). Then, an electronic MZI along a PN junction was realized by placing two valley beam splitters at both ends of the junction. However, quantum manipulation at the single-electron level, a crucial prerequisite for realizing electronic flying qubits, is still lacking in graphene. One primary reason is the absence of an electron pump, which generally requires dynamical control of quantum dots. This is extremely challenging in graphene because of the absence of a bandgap, unlike in conventional semiconductors.

In this work, we show that Levitons can be a reliable option for the injection of single electrons. Then, we demonstrate Bloch sphere rotation manipulations, taking advantage of the valley degrees of freedom in graphene MZI. This validates that the quantum coherence of the graphene MZI can be more than a few micrometers in length under high-frequency excitations. The information on final states is read statistically by combining conductance and noise measurements while periodically repeating the qubit operation.

On-demand injection of single electrons in graphene

In conventional semiconductors, different techniques for controlling on-demand electron excitation have been developed. The electron pump, composed of a series of tunnel barriers with an island or a dot, is one of them (*20*). Fast manipulation of the tunnel barriers enables sequential emission of electrons. In this case, electrons are injected well above the Fermi energy, which should limit Coulomb interactions with electrons from the Fermi [illegible]. However, because the electron excitations [illegible] far from the Fermi surface, there is more room to excite electron-hole pairs out of the ground state, leading to a very short relaxation length (*21*). For graphene, although the development of well-defined quantum dots is an active field (*22*), their fabrication remains challenging and an electron pump has not been demonstrated. An alternative approach to on-demand single-electron injection is a direct application of a voltage pulse $V(t)$, where t is time, on the emitter contact, with the condition that the Faraday flux $\varphi(t) = e\int V(t')dt'/h$ is an integer value, where e is the charge of the electron and h is Planck's constant. More specifically, it has been demonstrated that by shaping the pulse as a Lorentzian function, a single electron can be emitted without the creation of unwanted electron-hole pairs (*23*, *24*). This excitation has been called Leviton. In addition to its simplicity, this approach allows for the emission of electrons very close to the Fermi energy, where there is minimal room for electron-hole pair creation, thereby protecting the emitted electron from possible relaxation and decoherence. In the following sections, we present a demonstration of on-demand single-electron injection based on Levitons in graphene.

Coherent manipulation and reading of the graphene qubit

The next step was to demonstrate the coherent manipulation of single electrons during propagation. The most elementary quantum manipulation is the rotation of a single qubit on the Bloch sphere. This can be achieved through an electronic MZI, which can be formed in graphene by mixing two N- and P-type edge channels with opposite valley isospin $\pm\vec{w}$ in the bipolar quantum Hall regime (*10*). At the first electron beam splitter of the interferometer, the degree of valley-channel mixing can be characterized by a transmission probability T_1 and a reflection probability $R_1 = 1 - T_1$ of the beam splitter. The initial state of an electron defined after the first beam splitter can be written as a quantum-mechanical superposition $|\psi_{initial}\rangle = \cos\left(\frac{\theta_1}{2}\right)|\vec{w}\rangle + \sin\left(\frac{\theta_1}{2}\right)|-\vec{w}\rangle$, where we introduce the channel mixing angle θ_1 with $\cos\left(\frac{\theta_1}{2}\right) = r_1$ and $|r_1|^2 = 1 - T_1$, where r is a reflection coefficient. The valley superposition state evolves by acquiring the Aharonov-Bohm phase ϕ_{AB} (which is equal to $2\pi BA/\Phi_0$, where B is the magnetic field, A is the interferometer area, and $\Phi_0 = h/e$ is the flux quantum) along the MZI. The final state at arrival at the second beam splitter becomes $|\psi_{final}\rangle = \cos\left(\frac{\theta_1}{2}\right)|\vec{w}\rangle + \sin\left(\frac{\theta_1}{2}\right)e^{i\phi_{AB}}|-\vec{w}\rangle$. Therefore, beam splitters combined with the Aharonov-Bohm effect enable the basic operations of a valley-isospin flying qubit.

After the electron passes through the second beam splitter, the final state is measured by

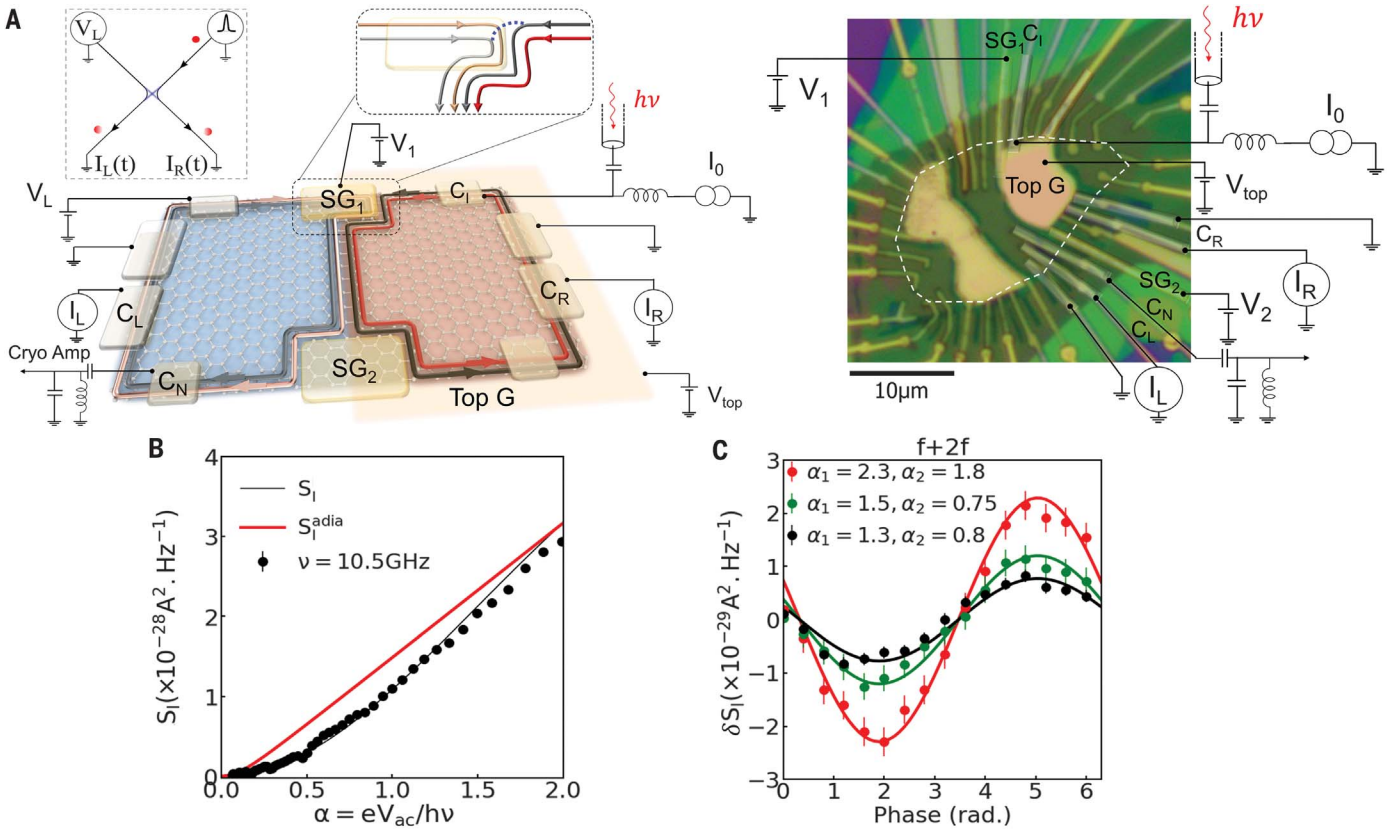


Fig. 1. Photoassisted shot noise in the quantum Hall regime. (**A**) (Left) Schematic representation of the device in the bipolar quantum Hall regime. The filling factor below the top gate (Top G) is tuned to $\nu_P = -2$, whereas the region not covered by the top gate is tuned to $\nu_N = 2$. The filling factors below the upper and lower side gates (SG_1 and SG_2, respectively) are tuned to $\nu_1 \geq 2$ and $\nu_2 = 0$, respectively. The voltage pulse at frequency υ is applied at the upper-right ohmic contact (C_I). Resulting photoassisted shot noise is amplified at low temperature (in contact C_N). Transmitted (I_L) and reflected current (I_R) are also measured. The inset on the left shows a schematic representation of the partitioning experiment. The inset on the right shows a zoomed-in view of the valley splitter, where edge-state mixing occurs. (Right) Microscopic view of the sample obtained through an optical microscope. (**B**) Measured shot noise S_I (black dots) as a function of the photon number α at 10.5 GHz. The data are compared with predicted photoassisted shot noise (black line) [see also section IC, "Engineering voltage pulses," in (*32*)] and the adiabatic (adia) excess noise (red line) given by $S_I^{adia} = 2e\frac{2e^2}{h}D(1-D)\frac{1}{T}\int_0^T dt V_{ac}\left|\sin\left(\frac{2\pi t}{T}\right)\right| = 2e\frac{2e^2}{h}D(1-D)\frac{2V_{ac}}{\pi}$. For $\alpha < 2$, the measured noise agrees with the predicted photoassisted shot noise. (**C**) Excess shot noise (circles) $\delta S_I = S_I(q = 1, \alpha_1, \alpha_2, \varphi) - S_I(q = -1, \alpha_1, \alpha_2, \varphi)$ as a function of the phase difference between the first and second harmonics for different amplitudes ($\alpha_1 = eV_1/h\upsilon$ and $\alpha_2 = eV_2/2h\upsilon$) compared with the predicted shot noise (solid lines) [see section IC, "Engineering voltage pulses," in (*32*)]. Error bars represent SEM.

projecting it on the output left state $|\tilde{\Psi}\rangle = \sin\left(\frac{\theta_2}{2}\right)|\vec{w}\rangle + \cos\left(\frac{\theta_2}{2}\right)|-\vec{w}\rangle$, with $\sin\left(\frac{\theta_2}{2}\right) = t_2$, where $T_2 = |t_2|^2$ represents the transmission probability of the second beam splitter. We measured the transmission probability of the MZI as $T_{MZI} = \left|\langle\tilde{\Psi}|\psi_{final}\rangle\right|^2$ and the associated noise. Using the noise measurement, we show that the fundamental property of a Leviton, namely the minimization of the number of electron-hole pairs, is preserved during its propagation through the interferometer.

Voltage-pulse generation

The sample we used is depicted in Fig. 1A. A global graphite back gate and a metallic top gate (labeled as Top G in Fig. 1A) are deposited on the right half of the sample in order to independently tune the electron density in the left and right halves of the sample. An electronic MZI can be formed in the bipolar quantum Hall regime (*10*, *18*, *19*, *25*–*28*). The filling factors of the N and P regions are set to $\nu_N = 2$ and $\nu_P = -2$, respectively, resulting in four copropagating channels along the PN junction. It has been shown that edge channels from the N and P regions with opposite valley polarization can be mixed by adjusting upper (SG_1) and lower (SG_2) side gates placed at the intersection between the PN junction and the physical edge of graphene (*10*).

We first consider the case where the valley beam splitter is formed only at the upper edge, which can be obtained by setting the filling factors below the upper and lower side gates to $\nu_1 \leq -2$ and $\nu_2 = 0$, respectively. We constructed periodic Lorentzian pulses by summing a series of harmonics with controlled amplitude and phase. Because the amplitude and phase of the pulse change during the propagation along the electromagnetic lines in the cryostat, the pulse emitted at the output of the generator at room temperature and the one that propagates along the lines of the cryostat into the sample at base temperature are different (Fig. 1A). To resolve this issue, it was crucial to determine the amplitude and phase of each harmonic that is required to build periodic Lorentzian pulses "in situ" by measuring the photoassisted shot noise (*7*). To this end, we first considered the simplest case with a single mode ($\upsilon = 10.5$ GHz). A sinusoidal potential at frequency υ is applied on the upper right ohmic contact (C_I), and the shot noise is measured at the lower left ohmic contact (C_N). By coherent scattering at the MZI, this generates a photoassisted shot noise, which is characterized by shot-noise singularities at $eV_{dc} = nh\upsilon$ (*29*–*31*), where V_{dc} is superimposed dc bias, n is an integer, and υ is frequency. The exact number of electron-hole pairs is computed by comparing measured photoassisted shot noise with established theoretical expectations (*6*, *29*) [see also section IIA, "Theoretical

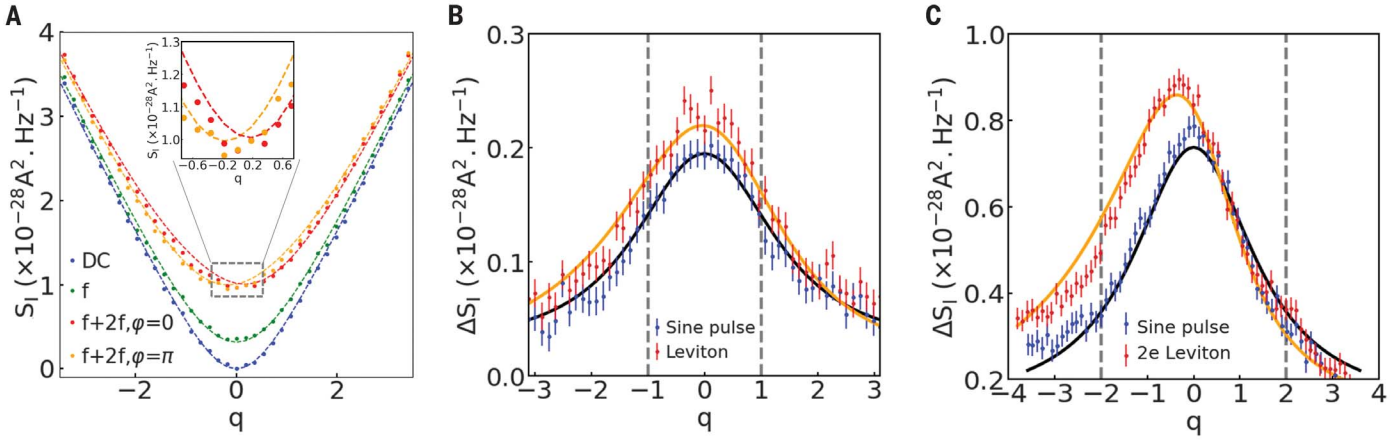


Fig. 2. Leviton and 2e Leviton. (**A**) Shot noise (circles) as a function of $q = eV_{dc}/h\nu$ for different values of α_1, α_2, and φ compared with the predicted shot noise (dashed lines) (*32*). Note that the finite temperature smears out the shot-noise singularities that are expected at $eV_{dc} = nh\upsilon$. The inset shows a zoomed-in view of the two-harmonics configuration. (**B**) Comparison of the shot-noise spectroscopy for a 3.5-GHz sine wave with amplitude $\alpha = 1$ (blue circles) and a 3.5-GHz Lorentzian pulse with $W/T = 0.09$ (W is the width of the pulse and $T = 1/f$, where f is the pulse repetition frequency) and $\alpha = 1$ (red circles). For the Lorentzian pulse, the excess shot noise ΔS_I is strongly asymmetric. For $q > \alpha + k_BT/h\upsilon$, the noise vanishes exponentially, which is the hallmark of the absence of hole excitations. Solid lines are theoretical predictions (*32*). (**C**) Same as (B) with $\alpha = 2$. Note that the offset $\Delta q = -0.22$ may result from the existence of photocurrent at $\alpha = 2$. We extracted an excess number of electron-hole pairs $\Delta N_{eh} = \Delta S_I/S_I^0 = 0.087$ for the Leviton and $\Delta N_{eh} = 0.076$ for the 2e Leviton. In (B) and (C), the orange and black lines are the photoassisted shot noise theory for Leviton and sine pulse, respectively, and error bars represent SEM.

model," in (*32*)]. The excess noise at the thermal equilibrium generated by the partitioning of electron-hole pairs at the first beam splitter can be expressed as $S_I = S_I^0 \sum_{l=-\infty}^{+\infty} l\left[\coth\left(\frac{l}{2\theta_e}\right) - 2\theta_e\right] J_l^2(\alpha)$, where $\theta_e = k_BT_e/h\upsilon$ is the temperature in frequency units, $\alpha = eV_{ac}/h\upsilon$ is the photon number, J_l is the Bessel function (l is an integer), and $S_I^0 = 2\frac{2e^2}{h}D(1-D)h\upsilon$ is the typical scale of the photoassisted shot noise (D is the transmission probability of one edge channel). Note that the presence of the factor 2 is caused by the two modes involved in the partitioning at the $\nu_N = 2$ and $\nu_P = -2$ configurations. The agreement with the theoretical photoassisted shot noise S_I confirms that the amplitude at the injection contact can be precisely determined (Fig. 1B). Next, to determine the phase, we measured the shot noise while varying the phase difference φ between the two harmonics, which defines the biharmonic signal $V_{ac,bi}(q, t, \varphi) = q + \alpha_1\cos(2\pi^*\upsilon t) + \alpha_2\cos(2\pi^*\upsilon t + \varphi)$, where $q = eV_{dc}/h\upsilon$, $\alpha_1 = eV_{ac,1}/h\upsilon$, and $\alpha_2 = eV_{ac,2}/2h\upsilon$. In Fig. 1C, $\delta S_I = S_I(q = 1, \alpha_1, \alpha_2, \varphi) - S_I(q = -1, \alpha_1, \alpha_2, \varphi)$ is plotted as a function of φ for $\upsilon = 3.5$ GHz. By comparing the results with the photoassisted theory (*32*), we could accurately adjust φ at the injection contact. We repeated this process for the third and fourth harmonics (see fig. S10 for a υ-3υ calibration). Combined with the amplitude control, this result demonstrates that it is possible to engineer any pulse shape at the injection contact [see also the detailed discussion in section IC, "Engineering voltage pulses," in (*32*)].

Having established that constructing a Lorentzian pulse is possible, we performed energy spectroscopy of the Fermi sea with the Lorentzian perturbation to demonstrate the minimization of electron-hole pair generation by Levitons. The idea was to apply a direct current (dc) on the upper-left ohmic contact, defining a voltage V_L, while the Lorentzian pulse is applied on the opposite right contact C_I. Under negative bias, electrons emitted in the energy range $-eV_L > \varepsilon > 0$, where ε is energy, will antibunch with electron excitations coming from the driven right contact, resulting in no noise (see also figs. S13 and S14). The noise variation with V_L gives a measure of the number of electron excitations. The same procedure was repeated with positive bias to extract the number of hole excitations. Figure 2A shows the photoassisted shot noise as a function of the dc bias for single and two-harmonic modes (at $\varphi = 0$ and $\varphi = \pi$), which agrees well with the theoretical prediction for any injected charge q, which is an important requirement for the injection of Leviton ($q = 1$) and 2e-Leviton ($q = 2$). In Fig. 2, B and C, the excess shot noise ΔS_I, obtained by subtracting the noise with V_{ac} "off" from the noise with V_{ac} "on," is shown for a Lorentzian pulse constructed by summing four harmonics with $\alpha = q = 1$ (Leviton configuration) and $\alpha = q = 2$ (2e-Leviton configuration). The asymmetry of the excess noise reflects the absence of hole creation by the Lorentzian pulse [see also "Excess

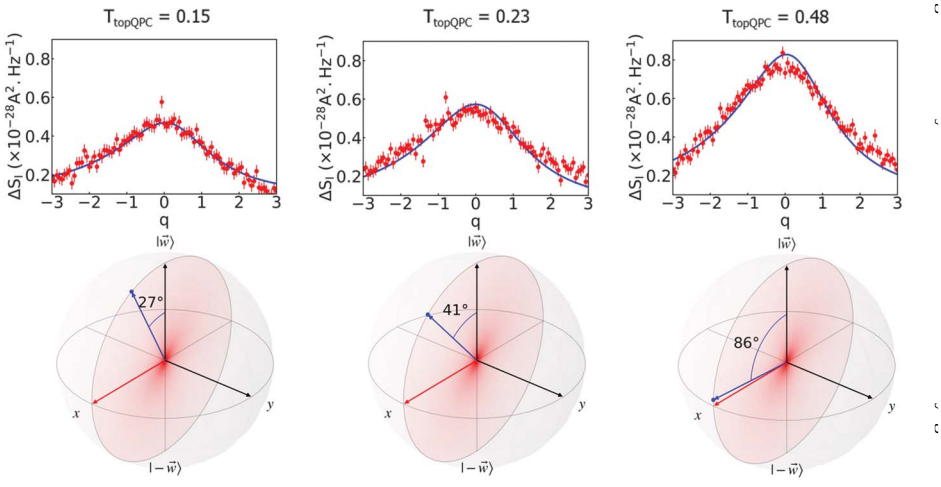


Fig. 3. Polar angle control of the 2e Leviton state. Excess shot-noise measurements at three different values of the transmission for the top beam splitter of the MZI and corresponding polar angles in the Bloch sphere representation. Error bars represent SEM. The solid blue line is the result of the photoassisted shot noise theory.

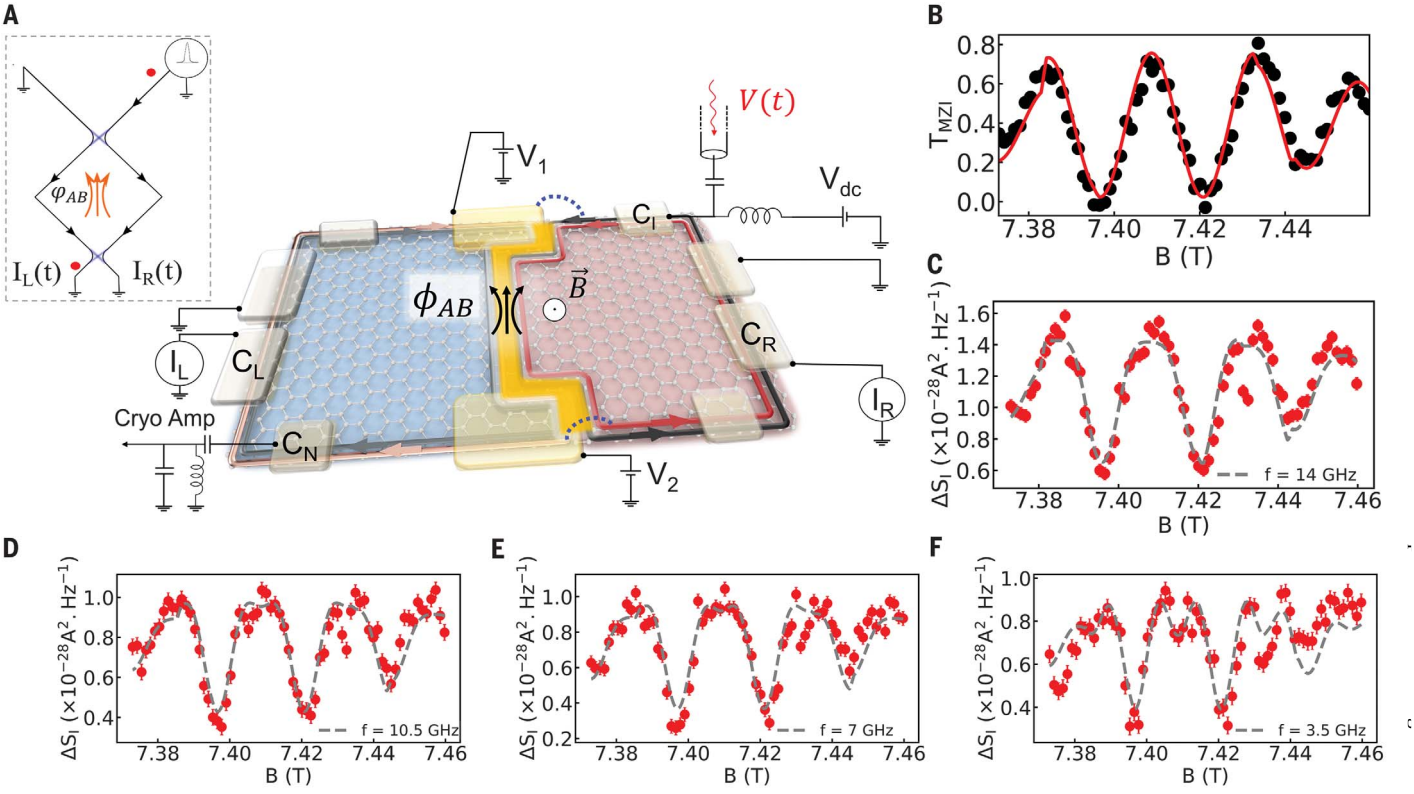


Fig. 4. On-demand excitations in MZI. (**A**) MZI configuration with side gates tuned to $\nu_1 \geq 2$ and $\nu_2 \geq 2$. The electron pulse now accumulates a phase ϕ_{AB} upon propagation (the area enclosed by the loop is shown in yellow). The inset shows a schematic representation of Levitons in a MZI experiment. (**B**) Interference oscillations of the MZI transmission T_{MZI} as a function of the magnetic field obtained by dc transport measurement. The solid red line is the theoretical prediction, taking into account decoherence as a free parameter (*10*). (**C**) Excess shot noise (red circles) generated by a sine wave propagating in a MZI as a function of the magnetic field for 14 GHz and $\alpha = 0.51$. The dashed gray line is the theoretical prediction. [Detailed equations are given in "Floquet scattering formalism for graphene Mach-Zehnder interferometry" in section IIA of (*32*).] (**D** to **F**) As in (C) but for 10.5 GHz ($\alpha = 0.55$) (D), $f = 7$ GHz ($\alpha = 0.64$) (E), and $f = 3.5$ GHz ($\alpha = 1.22$) (F).

noise from sine and Lorentzian excitations" in section IC of (*32*)]. By comparing it to the computed excess noise for an ideal Lorentzian pulse (solid orange line), we experimentally verified that one or two electrons are injected along with a minimum amount of electron-hole pairs (at finite temperature thermal excitations add an extra contribution to the excess noise). This constitutes an experimental demonstration of on-demand single-electron injection in graphene quantum Hall edge channels. Note that for one- and two-electron Levitons, we did not observe any deviation from the non-interacting theory [see also "Relevance of the electronic interactions" in section ID of (*32*)]. This measurement can be realized at different values of the transmission, which demonstrates the polar angle θ_1 control of the 2e-Leviton state (Fig. 3).

Coherent manipulation on the Bloch sphere

We then turned on the lower beam splitter to showcase the coherent manipulation of single electrons as they propagate through the electronic MZI (Fig. 4, A and B). After passing the first beam splitter located at the upper edge of the PN junction, excited single electrons propagate along the N side with transmission probability $T_1 = |t_1|^2$ or the P side with reflection probability $|r_1|^2 = 1 - T_1$. In the Bloch sphere representation, the polar angle of the valley-isospin qubit is tuned by the upper side gate. When the polar angle is chosen to be $\frac{\pi}{2}$ (where the upper beam splitter is half-open as $T_1 = |t_1|^2 = 0.5$), the valley isospin of the initial state $|\Psi_{initial}\rangle = \left(\left|\vec{w}\right\rangle + \left|-\vec{w}\right\rangle\right)/\sqrt{2}$ at the entrance of the PN junction lies on the equator. The isospin then rotates around the z axis by the azimuthal angle ϕ_{AB}. The final state at the lower edge of the PN junction becomes $|\Psi_{final}\rangle = \left(\left|\vec{w}\right\rangle + e^{i\phi_{AB}}\left|-\vec{w}\right\rangle\right)/\sqrt{2}$. The value of ϕ_{AB} can be measured by the transmission probability $T_{MZI} = \left|\langle\tilde{\Psi}|\psi_{final}\rangle\right|^2$, whereas the number of electron-hole pairs at the lower beam splitter can be measured by the shot noise.

We first applied this operation to sine-shaped voltage pulses. The shot noise is essentially determined by the total number of electron-hole pairs created by the pulse multiplied by the factor $T_{MZI}(1 - T_{MZI})$. It is given by two-particle scattering processes that enclose the AB-flux once (leading to a phase contribution $\phi_{AB} = 2\pi BA/\Phi_0$) or two times ($2\phi_{AB}$) (*33*, *34*) [see also "Floquet scattering formalism for graphene Mach-Zehnder interferometry" in section IIA of (*32*) for the full formula]. Figure 4, C to F, shows the shot noise as a function of the magnetic field for several frequency values ($\upsilon = 3.5$, 7, 10.5, and 14 GHz), which are used to construct Levitons. The oscillation of the shot noise with ϕ_{AB} indicates that the quantum coherence of the MZI survives under high-frequency excitations.

Finally, to demonstrate the rotation of an electron quantum state on the equator of the Bloch sphere, it is necessary to measure the number of electron-hole pairs while Levitons are injected periodically and ϕ_{AB} is allowed to evolve. We evaluated $\Delta S_I(q)$ generated by periodic 2e-Leviton injection for different values of the magnetic field (Fig. 5, A and B). Note that the signal-to-noise ratio for 1e-Leviton excitation was too low to perform a precise measurement. The observed asymmetry of $\Delta S_I(q)$ measured at the output of the MZI for all the

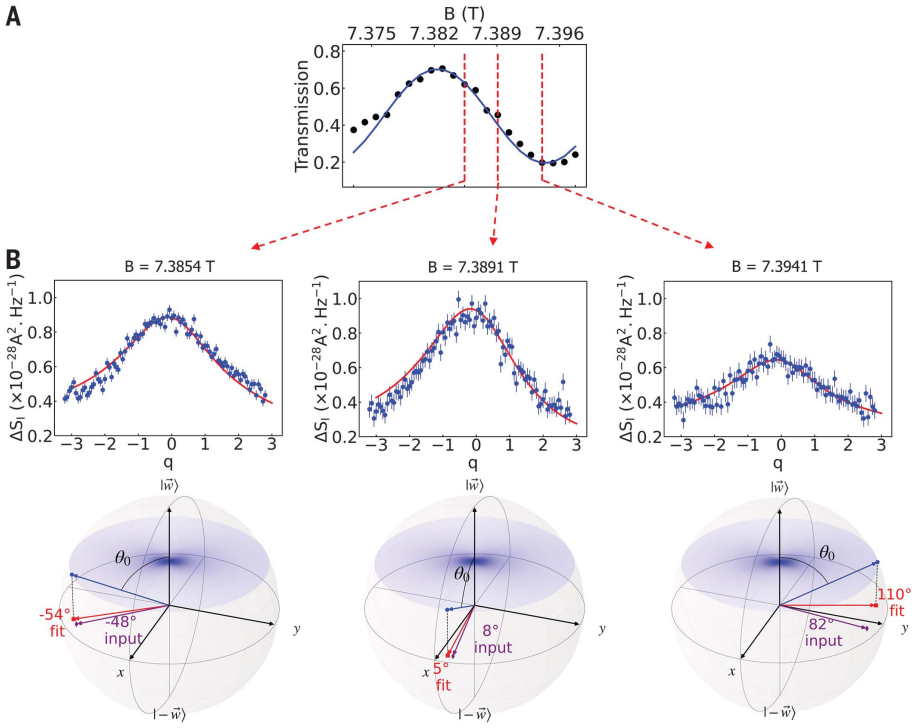


Fig. 5. Coherent manipulation of a 2e Leviton. (**A**) MZI transmission T_{MZI}. The solid blue line is a sinusoidal fit used to extract the ϕ_{AB}. (**B**) (Top) Excess shot noise generated by a Leviton propagating in a MZI interferometer for different values of ϕ_{AB} (B = 7.3854, 7.3891, and 7.3941 T). (Bottom) From left to right, the valley isospin rotates by an azimuthal angle of ~π on the Bloch sphere. Fitted ϕ_{AB} (labeled "fit") is compared to the actual value of ϕ_{AB} set by the magnetic field (labeled "input"). The solid red line is fit to the photoassisted shot noise theory. Error bars represent SEM.

investigated ϕ_{AB} values verifies that the key property of Leviton, namely the minimization of the number of electron-hole pairs with 2e injected charges, is conserved while propagating in the MZI. Furthermore, the value of ϕ_{AB} that is extracted from the amplitude of $\Delta S_I(q)$ is consistent with that obtained from T_{MZI}. These results demonstrate coherent control of Levitons.

Conclusions and outlook

Our study demonstrates the emission and coherent control of a quantum state at the single-electron level in monolayer graphene. Although graphene interferometers have been studied in the dc regime (*10*, *18*, *19*, *35*, *36*), sending on-demand excitations or flying qubits toward a MZI is a notable step toward quantum information transfer. We established an on-demand electron source in graphene based on Levitons that minimizes the number of unwanted electron-hole pairs. By sending periodically excited Levitons to the MZI, we demonstrated the rotation of the valley flying qubit on the Bloch sphere. Encoding quantum information in valley state of Levitons should enable two-valley qubit operations to be considered (*37*–*41*). There it can be shown that minimizing the number of electron-hole pairs is relevant against decoherence [see also "Levitons and decoherence" in section ID of (*32*)]. These on-demand electron pulses can also carry fractional charges, which offers the possibility to braid anyons (*42*–*46*) in graphene in the time domain. Graphene is emerging as a promising material with robust quantum properties compared with those of conventional semiconductors. Beam splitters, interferometers, and single-electron sources can be easily realized using PN junctions, opening up avenues for electronic quantum optics experiments. Owing to the simple and elegant circuit topology that exploits N and P counter-propagating edge states, complex, yet compact, interferometers with original entanglement schemes can be envisioned.

REFERENCES AND NOTES

1. C. Bäuerle *et al*., *Rep. Prog. Phys.* **81**, 056503 (2018).
2. Y. Ji *et al*., *Nature* **422**, 415–418 (2003).
3. P. Roulleau *et al*., *Phys. Rev. Lett.* **100**, 126802 (2008).
4. M. Yamamoto *et al*., *Nat. Nanotechnol.* **7**, 247–251 (2012).
5. E. Bocquillon *et al*., *Science* **339**, 1054–1057 (2013).
6. J. Dubois *et al*., *Nature* **502**, 659–663 (2013).
7. D. C. Glattli, P. Roulleau, *Phys. Status Solidi, B Basic Res.* **254**, 1600650 (2017).
8. T. Jullien *et al*., *Nature* **514**, 603–607 (2014).
9. D. Dasenbrook, C. Flindt, *Phys. Rev. B* **92**, 161412 (2015).
10. M. Jo *et al*., *Phys. Rev. Lett.* **126**, 146803 (2021).
11. M. Jo *et al*., *Nat. Commun.* **13**, 5473 (2022).
12. J. R. Schaibley *et al*., *Nat. Rev. Mater.* **1**, 16055 (2016).
13. K. F. Mak, K. L. McGill, J. Park, P. L. McEuen, *Science* **344**, 1489–1492 (2014).
14. Y. Shimazaki *et al*., *Nat. Phys.* **11**, 1032–1036 (2015).
15. L. Ju *et al*., *Nature* **520**, 650–655 (2015).
16. J. Li *et al*., *Science* **362**, 1149–1152 (2018).
17. R. V. Gorbachev *et al*., *Science* **346**, 448–451 (2014).
18. S. Morikawa *et al*., *Appl. Phys. Lett.* **106**, 183101 (2015).
19. D. S. Wei *et al*., *Sci. Adv.* **3**, e1700600 (2017).
20. L. P. Kouwenhoven, A. T. Johnson, N. C. van der Vaart, C. J. P. M. Harmans, C. T. Foxon, *Phys. Rev. Lett.* **67**, 1626–1629 (1991).
21. R. H. Rodriguez *et al*., *Nat. Commun.* **11**, 2426 (2020).
22. M. Eich *et al*., *Phys. Rev. X* **8**, 031023 (2018).
23. D. A. Ivanov, H. W. Lee, L. S. Levitov, *Phys. Rev. B* **56**, 6839–6850 (1997).
24. J. Keeling, I. Klich, L. S. Levitov, *Phys. Rev. Lett.* **97**, 116403 (2006).
25. C. Handschin *et al*., *Nano Lett.* **17**, 5389–5393 (2017).
26. P. Makk *et al*., *Phys. Rev. B* **98**, 035413 (2018).
27. J. Tworzydło, I. Snyman, A. R. Akhmerov, C. W. J. Beenakker, *Phys. Rev. B* **76**, 035411 (2007).
28. L. Trifunovic, P. W. Brouwer, *Phys. Rev. B* **99**, 205431 (2019).
29. L.-H. Reydellet, P. Roche, D. C. Glattli, B. Etienne, Y. Jin, *Phys. Rev. Lett.* **90**, 176803 (2003).
30. M. Kapfer *et al*., *Science* **363**, 846–849 (2019).
31. J. Gabelli, B. Reulet, *Phys. Rev. B* **87**, 075403 (2013).
32. See supplementary materials.
33. P. P. Hofer, C. Flindt, *Phys. Rev. B* **90**, 235416 (2014).
34. F. Battista, F. Haupt, J. Splettstoesser, *Phys. Rev. B* **90**, 085418 (2014).
35. C. Déprez *et al*., *Nat. Nanotechnol.* **16**, 555–562 (2021).
36. Y. Ronen *et al*., *Nat. Nanotechnol.* **16**, 563–569 (2021).
37. I. Neder *et al*., *Nature* **448**, 333–337 (2007).
38. P. Samuelsson, E. V. Sukhorukov, M. Büttiker, *Phys. Rev. Lett.* **92**, 026805 (2004).
39. N. Ubbelohde *et al*., *Nat. Nanotechnol.* **18**, 733–740 (2023).
40. J. Wang *et al*., *Nat. Nanotechnol.* **18**, 721–726 (2023).
41. J. D. Fletcher *et al*., *Nat. Nanotechnol.* **18**, 727–732 (2023).
42. J. Nakamura, S. Liang, G. C. Gardner, M. Manfra, *Nat. Phys.* **16**, 931–936 (2020).
43. H. Bartolomei *et al*., *Science* **368**, 173–177 (2020).
44. J. M. Lee, C. Han, H.-S. Sim, *Phys. Rev. Lett.* **125**, 196802 (2020).
45. M. P. Rösli *et al*., *Sci. Adv.* **7**, eabf5547 (2021).
46. J. M. Lee *et al*., *Nature* **617**, 277–281 (2023).
47. A. Assouline *et al*., Emission and coherent control of Levitons in graphene. Zenodo (2023); https://zenodo.org/records/10044265.

ACKNOWLEDGMENTS

We thank W. Dumnernpanich for his help with fabrication. **Funding:** This work was funded by European Research Council (ERC) starting grant COHEGRAPH 679531 (P.R.); by European Metrology Programme for Innovation and Research (EMPIR) project SEQUOIA 17FUN04, which is cofinanced by the participating states and the European Union's Horizon 2020 program (P.R.); by the National Research Foundation of Korea via the SRC Center for Quantum Coherence in Condensed Matter (grant no. 2016R1A5A1008184 and RS-2023-00207732) (H.-S.S.); and by "Investissements d'Avenir" LabEx PALM (ANR-10-LABX-0039-PALM) (Project ZerHall) (F.D.P.). **Author contributions:** A.A., L.P., H.C., L.B., and P.R. performed the experiment with help from F.D.P.; A.A., L.P., H.C., L.B., N.K., S.L., H.-S.S., D.C.G., F.D.P., and P.R. analyzed and discussed the data; T.T. and K.W. provided the boron nitride layers; M.J. fabricated the device with input from A.A., F.D.P., and P.R.; P.R. wrote the manuscript with input from all co-authors; and P.R. designed and supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data and code are archived at Zenodo (*47*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf9887
Materials and Methods
Supplementary text
Figs. S1 to S20
References (*48*–*56*)

Submitted 24 November 2022; accepted 8 November 2023
10.1126/science.adf9887

THERMAL TRANSPORT

Low voltage–driven high-performance thermal switching in antiferroelectric $PbZrO_3$ thin films

Chenhan Liu[1]†, Yangyang Si[2]†, Hua Zhang[3]†, Chao Wu[3], Shiqing Deng[4], Yongqi Dong[5], Yijie Li[2], Meng Zhuo[1], Ningbo Fan[6], Bin Xu[6], Ping Lu[1], Lifa Zhang[7], Xi Lin[2], Xingjun Liu[2], Juekuan Yang[3], Zhenlin Luo[5], Sujit Das[8], Laurent Bellaiche[9], Yunfei Chen[3]*, Zuhuang Chen[2,10]*

Effective control of heat transfer is vital for energy saving and carbon emission reduction. In contrast to achievements in electrical conduction, active control of heat transfer is much more challenging. Ferroelectrics are promising candidates for thermal switching as a result of their tunable domain structures. However, switching ratios in ferroelectrics are low (<1.2). We report that high-quality antiferroelectric $PbZrO_3$ epitaxial thin films exhibit high-contrast (>2.2), fast-speed (<150 nanoseconds), and long-lifetime (>10^7) thermal switching under a small voltage (<10 V). In situ reciprocal space mapping and atomistic modelings reveal that the field-driven antiferroelectric-ferroelectric phase transition induces a substantial change of primitive cell size, which modulates phonon-phonon scattering phase space drastically and results in high switching ratio. These results advance the concept of thermal transport control in ferroic materials.

The recovery and management of waste heat is important for energy saving and carbon emission reduction (*1*–*3*) as ~60% of the energy in the world is wasted as heat. This requires the effective utilization of waste heat and active control of heat transfer (*4*–*6*). One form of heat transfer is phonon transport (*7*–*9*). Although modulating phonon transport poses greater challenges compared with electron transport, thanks to the conceptualization of phononic devices (*6*), active and reversible control of phonon transport to mitigate heat-related concerns has stimulated broad research interest (*5*, *10*, *11*). Until now, active and reversible control of phonon transport or thermal conductivity k (i.e., thermal switching) has been explored by many different methods: strain engineering (*10*, *12*), light-triggered alignment of molecular chains (*13*, *14*), electrochemical-controlled phase transition (*15*) or ion intercalation (*16*–*18*), temperature-triggered phase transition (*19*, *20*), and electric field–driven domain wall density change (*21*–*23*), etc. For practical applications, high-performance thermal switching should meet three crucial conditions: (i) high switching ratio, (ii) large number of switching cycles, and (iii) short switching time. Despite extensive research on thermal switching methods in various materials, it has been difficult to find a material that fulfills all three essential conditions at room temperature. For example, the electrochemical intercalation of atoms or ions into layered materials can induce a substantial change in thermal conductivity (*16*–*18*). However, the intercalation process needs a long time (tens of minutes to several hours) and the subsequent switching ratio severely degenerates after only a few cycles (*15*).

Ferroelectric (FE) materials are potential candidates for thermal switching because they have fast dipole switching (*24*) and electrically tunable domain structures (*25*–*28*), which can be employed to modulate thermal conductivity (*21*–*23*) by the effects of phonon-domain wall scattering (*29*–*31*). However, the observed switching ratios ($\lambda = k_{on}/k_{off}$) in ferroelectrics are rather low, around 1.1 to 1.2 (*21*–*23*, *32*, *33*), at room temperature. Recent observations have indicated that the low switching ratios originate from a mismatch between domain size and phonon mean free path (*32*). That is, near room temperature the Umklapp-dominated phonon mean free path (in the range of a few nanometers to tens of nanometers) (*32*, *34*, *35*) is much shorter than the domain size (in the range of tens to hundreds of nanometers) (*23*, *32*), resulting in the weak effects of domain walls on thermal conductivity. Because a phase transition can change the lattice structure directly, which may achieve wide-range modulation of phonon mean free path, it can also be employed for thermal switching in ferroic materials (*34*, *36*, *37*). However, observations report phase transition–induced thermal switching driven by an electric field are absent. We investigated the potential for high-performance thermal switching in antiferroelectrics (AFEs) (*38*, *39*) by exploring electric field–driven thermal switching through a structural phase transition.

Using prototypical AFE $PbZrO_3$ (PZO) as a model system, we demonstrated that electric field–driven AFE-FE phase transition has the potential for high λ because of the substantial difference between the AFE and FE structures. Through careful control of the film thickness and orientation, we achieved a fourfold modulation of the number (n) of atoms in the primitive cell and observed a 2.2-fold switching of thermal conductivity at room temperature. This switching ratio is twice as high as previously reported ratios in FE materials. In addition, our results demonstrated that the applied voltage required to induce the phase transition is less than 10 V, the switching time is less than 150 ns, and the number of switching cycles exceeds 10^7.

Mechanism for thermal switching

We employed electric field–triggered AFE-FE phase transition to realize the thermal switching. We show a schematic diagram for the reversible modulation of thermal conductivity through the phase transition in Fig. 1. At zero external field, the structure of PZO is AFE with space group of *Pbam*, which has "↑↑↓↓" antiparallel dipole arrangement (*38*). In this case, the structure is complex and n has a high value of 40, leading to a low k, i.e., the "OFF" state. When a sufficiently strong electric field is applied, the structure will transform from the AFE to FE phase, causing the decrease of structural complexity and n. As a result, k increases and the state changes to "ON". Crucially, upon removing the external electric field, the structure of PZO returns to its original AFE phase as a result of its lower energy state, thereby achieving reversible AFE-FE phase transition and thermal switching.

Strategies to improve switching ratio

For thermal switching, switching ratio λ is the key factor to evaluate its performance. From the above discussion, to obtain the highest switching ratio in PZO films, two conditions should be satisfied, i.e., the lowest thermal conductivity k_{off} at the "OFF" state and the highest thermal conductivity k_{on} at the "ON" state. For the first condition, the AFE structure at zero field should be pure with small remnant polarization P_r since the FE component might easily exist in PZO films (*40*, *41*) and thus increase k_{off}. For the second condition, a wider-range modulation of n is beneficial since smaller n in the FE structure can lead to a higher k_{on}.

Previous studies have demonstrated that P_r of PZO films exhibits an increasing trend as

[1]Micro- and Nano-scale Thermal Measurement and Thermal Management Laboratory, Jiangsu Key Laboratory for Numerical Simulation of Large-Scale Complex Systems, School of Energy and Mechanical Engineering, Nanjing Normal University, Nanjing 210023, P. R. China. [2]School of Materials Science and Engineering, Harbin Institute of Technology, Shenzhen, Guangdong 518055, P. R. China. [3]Jiangsu Key Laboratory for Design and Manufacture of Micro-Nano Biomedical Instruments, School of Mechanical Engineering, Southeast University, Nanjing 211100, P. R. China. [4]Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, Beijing 100083, P. R. China. [5]National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei, Anhui 230026, P. R. China. [6]Institute of Theoretical and Applied Physics, Jiangsu Key Laboratory of Thin Films, School of Physical Science and Technology, Soochow University, Suzhou, Jiangsu 215006, P. R. China. [7]Phonon Engineering Research Center, School of Physics and Technology, Nanjing Normal University, Nanjing 210023, P. R. China. [8]Material Research Centre, Indian Institute of Science, Bangalore 560012, India. [9]Physics Department and Institute for Nanoscience and Engineering, University of Arkansas, Fayetteville, AR 72701, USA. [10]Flexible Printed Electronics Technology Center, Harbin Institute of Technology, Shenzhen, Guangdong 518055, P.R. China.
*Corresponding authors. Email: zuhuang@hit.edu.cn (Z.C.); yunfeichen@seu.edu.cn (Y.C.)
†These authors contributed equally to this work.

the film thickness decreases, suggesting the coexistence of a metastable FE phase when the film is thin (<100 nm) (*42*–*44*). Thus, to mitigate the impact of the FE component we epitaxially grew high-quality PZO films with different thicknesses on $SrRuO_3$ (SRO) buffered $SrTiO_3$ (STO) substrates. We found that the optimal thickness was 150 nm (fig. S1). Furthermore, we employed STO substrates with different orientations to explore the response of antiferroelectric anisotropy to an external electric field for thermal switching. The reason is that the lattice structure and electric response of PZO films would be different as the electric field is applied along different directions (*45*).

Our x-ray diffraction θ-2θ scans (Fig. 2A) show good correspondence among diffraction peaks of (100)-, (110)-, and (111)-oriented PZO films and we found no secondary impurity phases. A representative atomic force microscopy topographic image (inset of Fig. 2A, fig. S2, and supplementary note 1) shows a smooth film surface with a root mean square roughness of 192 pm, indicating high quality of films and interfaces (fig. S3). The double hysteresis loops of the three oriented PZO films (Fig. 2B) show the near-zero remnant polarization (P_{r+} and P_{r-}), indicating the almost pure antiferroelectricity of 150-nm-thick films. Notably, the (111)-oriented PZO film has the smallest P_r (as low as ~2 μC/cm^2) and the largest saturation polarization P_s (as high as ~60 μC/cm^2), compared to the (110)- and (100)-oriented PZO films, indicating the presence of the smallest FE component at a zero electric field and the simplest structure at a sufficiently strong electric field. Thus, by controlling film growth thickness and orientation, a low k_{off} and a high k_{on} could be achieved.

To reveal the AFE features at zero field, we further analyzed the structure of (111)-oriented PZO heterostructure. By employing the x-ray diffraction reciprocal space mapping (RSM) around (312)-reflection of STO (Fig. 2C), 1/4 satellite peaks were observed along the in-plane direction, indicating the presence of ordered arrangement of AFE dipoles. Furthermore, to obtain atomic-scale information regarding the AFE feature, we conducted high-angle annular dark-field (HAADF) scanning transmission electron microscopy (STEM) imaging (Fig. 2D). We confirmed the antipolar nature ↑↑↓↓ of Pb^{2+} displacement within the selected area in Fig. 2D (dashed rectangle) by calculating the Pb^{2+} shift mapping using its position deviation from the averaged position of

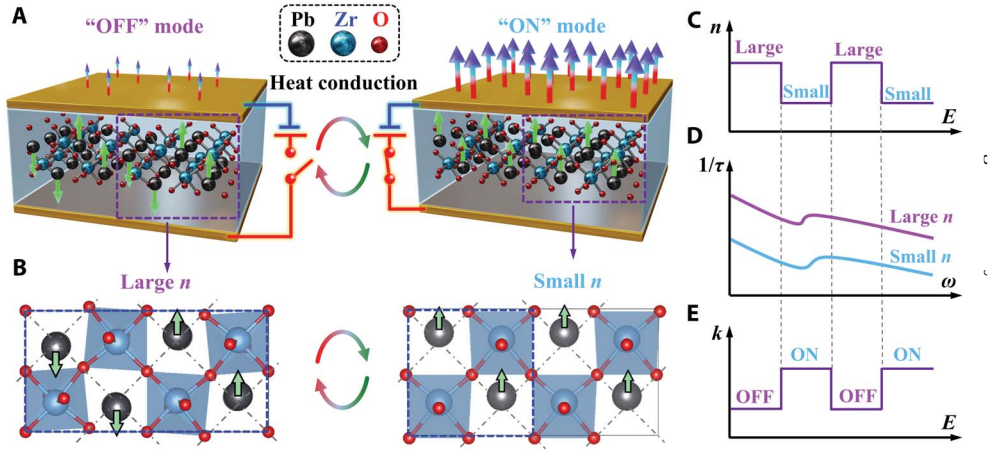


Fig. 1. Unlocking the mechanism for reversible modulation of *n* and thermal conductivity. (**A**) Schematic diagram of external voltage- or electric field–triggered "ON" and "OFF" modes. (**B**) Schematic diagram for the reversible modulation of *n*. At the "OFF" state, PZO has an AFE phase with a large *n*, whereas at the electric field–triggered "ON" state, PZO has an FE phase with a small *n*. The dashed blue rectangle represents the primitive unit cell, which has a significant change in the AFE-FE phase transition. (**C**) Electric field–controlled *n*. (**D**) Phonon frequency ω–dependent phonon-phonon scattering rate (reciprocal of phonon lifetime τ). Larger *n* has a higher scattering rate or shorter phonon lifetime while smaller *n* has lower scattering rate or longer phonon lifetime. (**E**) Electric field–controlled "ON" and "OFF" modes.

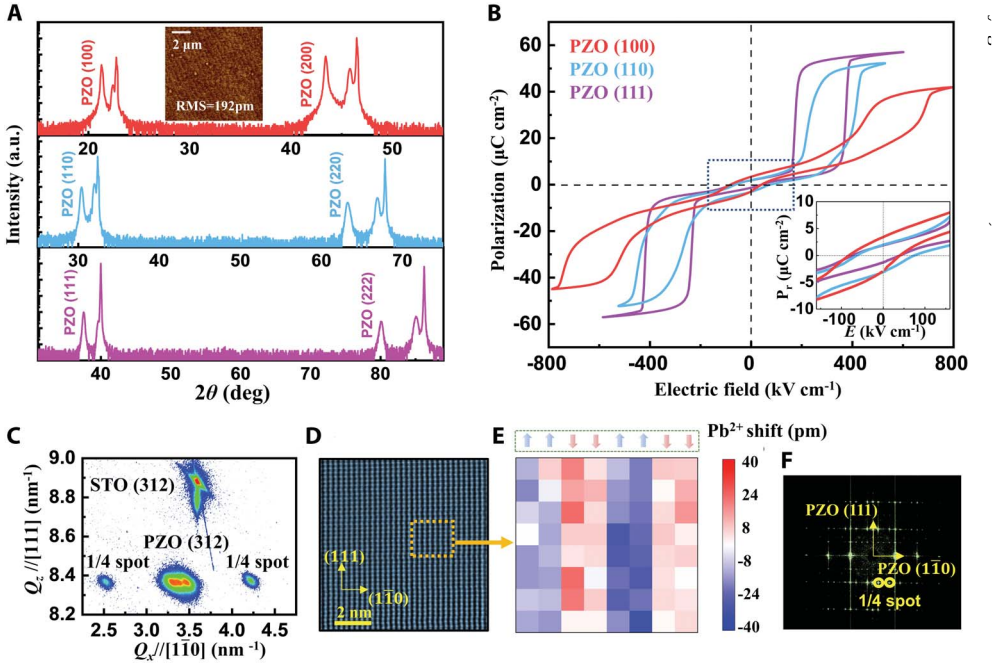


Fig. 2. Structural characterization of single-crystal PZO. (**A**) θ-2θ scans of PZO/SRO heterostructures on STO (100), (110), and (111) substrates. The inset shows the typical surface topography characterized by atomic force microscopy (AFM). (**B**) Double hysteresis loops of (111)-, (110)-, and (100)-oriented PZO thin films. The inset presents a magnified region of the doted square for the comparison of remnant polarization P_r. (**C**) Reciprocal space mapping (RSM) around STO (312)-diffraction condition for the (111) $PbZrO_3$ heterostructure. (**D**) A typical HAADF-STEM cross-sectional image of the (111) film. (**E**) Heat mapping of Pb^{2+} shift in the selected area in (D) (dashed rectangle), which clearly shows the AFE structure. (**F**) Fast Fourier transform from (D), revealing the 1/4 antiferroelectric modulation marked by yellow circles.

two adjacent Zr^{4+} ions along the [111] direction, as depicted in Fig. 2E. The presence of 1/4 superstructure spots in the fast Fourier transform of the STEM image provides further evidence for the fourfold periodicity of the atomic arrays, confirming the atomic-scale AFE order (Fig. 2F). This observation establishes that the pristine PZO films under investigation exhibit an excellent antiferroelectric feature in both reciprocal and real spaces.

In situ measurement of thermal switching

To experimentally verify our analysis, we performed in situ measurement of electric field–triggered thermal switching in PZO films through use of the time-domain thermoreflectance (TDTR) technique (fig. S4). To facilitate the in situ measurement of k under an external field, we designed a capacitor heterostructure Pt/PZO/SRO (fig. S3). We show the electric field–dependent thermal conductivities of (100)-, (110)-, and (111)-oriented PZO films at room temperature (Fig. 3A). Compared with other oriented films the (111)-oriented film exhibits similarly low k at zero field but demonstrates higher k under the electric fields ranging from −800 to 800 kV/cm. This observation suggests that the (111)-oriented film possesses the highest switching ratio. We show the $-V_{in}/V_{out}$ signals from the TDTR measurements for the three oriented PZO films at an electric field of −600 kV/cm (−9 V) (Fig. 3B), at which the AFE-FE phase transition (Fig. 2B) is induced. We obtained k of the film by fitting the measured signal (solid lines, Fig. 3B). We found that the (111)-oriented PZO film has the fastest $-V_{in}/V_{out}$ decay versus delay time and accordingly has the largest k (1.28 W m^{-1} K^{-1}), whereas the (100)-oriented PZO film has the slowest drop and lowest k (0.89 W m^{-1} K^{-1}). We compare the switching ratios of the three PZO films in Fig. 3D. The (111)-oriented PZO film demonstrates a switching ratio of 2.2, indicating that the electric field can induce the maximum modulation range of n in this crystallographic orientation. Our finding is consistent with the polarization measurement results and our first-principles modeling (Fig. 4).

For practical applications, thermal switch devices require a long lifetime which typically necessitates a substantial number of switching cycles. However, measuring one thermal cycle (Fig. 3A) takes several hours, making it unrealistic to measure a large number of thermal cycles. Because thermal switching in PZO is determined by the field-driven AFE-FE phase transition, the lifetime of PZO-based thermal switch devices can be indirectly determined by measuring the number of AFE-FE phase transitions, which corresponds to the number of electric cycles. We show the cycle number–dependent thermal switching in a (111)-oriented PZO film, which can be divided into three steps: (i) We directly measured the thermal conductivity for the first eight cycles, which shows excellent reversibility; (ii) we performed 10^7 electric cycles following the thermal measurements. We did not observe discernible electric fatigue, and large saturation polarization persists after 10^7 cycles. Our observations suggest that the reversible AFE-FE phase transition exhibits essentially no degradation, which is further supported by the hysteresis loop (fig. S6A); (iii) we conducted additional four thermal cycles after 10^7 electric cycles and the switching ratio remained comparable to the initial value. Our observation suggests that PZO-based thermal switch devices have the potential for a large number of thermal switching cycles, likely resulting in a long lifetime. We attribute the long-lifetime performance to the superior quality of the films (*46*). To confirm that damage reduces the switching ratio, we also measured the thermal conductivity of a fatigued sample after 10^8 electric cycles. We found that this sample exhibited a substantial increase in k compared to the as-grown state (fig. S7B), corresponding to a substantial reduction in the switching ratio. The hysteresis loop analysis reveals that the

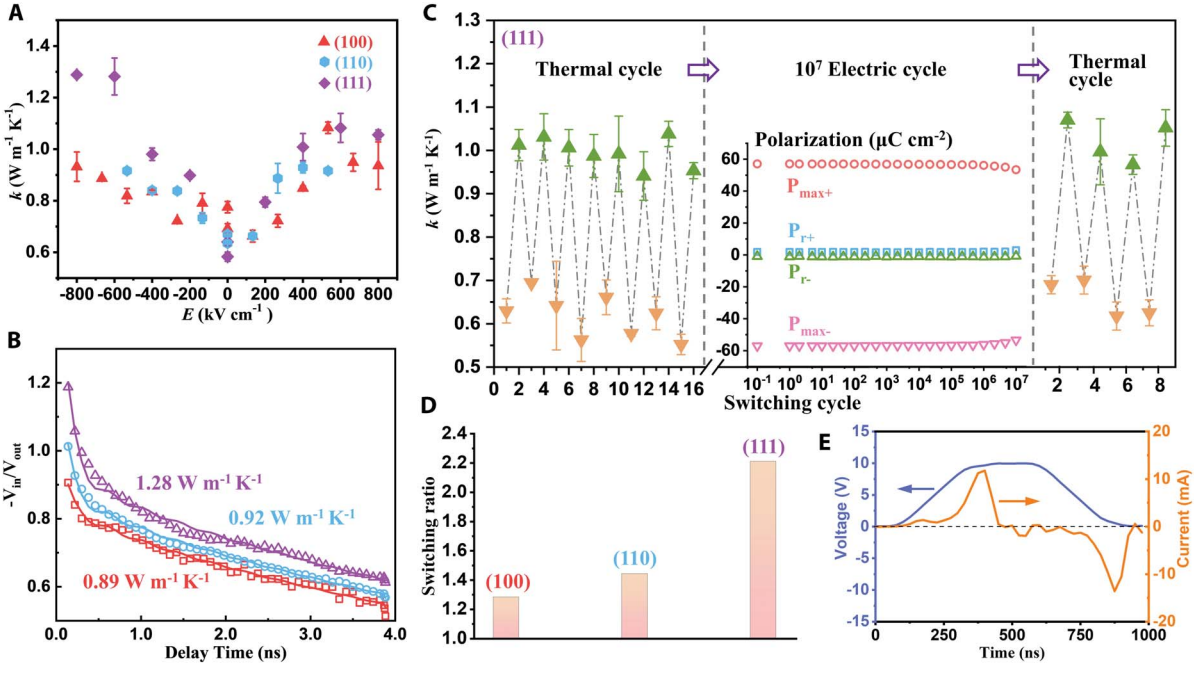


Fig. 3. Electric field–triggered thermal switching in PZO at room temperature. (**A**) Electric field–dependent thermal conductivity k of (111)-, (110)-, and (100)-oriented PZO films. The thermal conductivity has a significant increase independent on the sign (negative or positive) of the electric field and the largest thermal conductivity appears at E higher than ±600 kV/cm, i.e., ±9 V. (**B**) The comparison of $-V_{in}/V_{out}$ signals for (111)-, (110)-, and (100)-oriented PZO films at an electric field around −600 kV/cm. (**C**) Switching cycle–dependent thermal conductivity and polarization for (111)-oriented PZO film: 8 thermal cycles, 10^7 electric cycles, and 4 final thermal cycles. (**D**) The comparison of switching ratios for PZO films at an electric field around −600 kV/cm. Another comparison of switching ratios at an electric field of 400 kV/cm is shown in fig. S5. (**E**) Time-dependent switching current (orange) and voltage (light purple). From the current profile, the switching time is extracted to be about 150 ns. Each point in (A) and (C) is representative of at least three independent experiments, each with similar results. Bar, mean ± standard error.

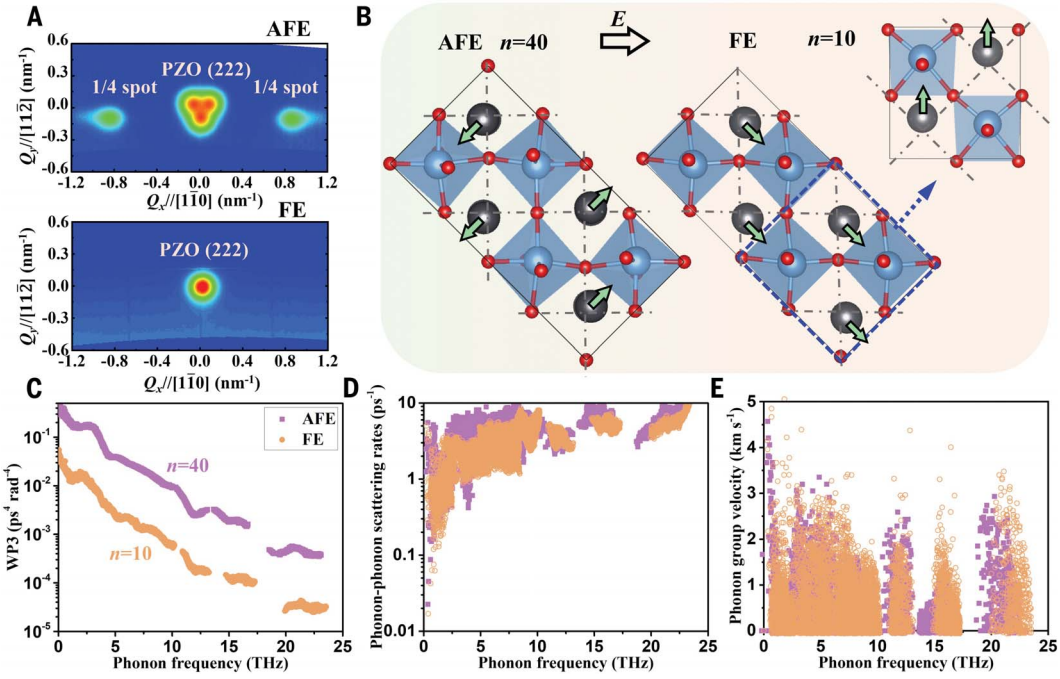


Fig. 4. Atomic-scale analysis of thermal switching in PZO. (**A**) In-plane RSM from the 3D synchrotron–based XRD data in AFE (upper panel) and FE (lower panel) phases. The disppearance of 1/4 satellite peaks and the merging of split main Bragg peaks as the structure transforms from AFE to FE indicate the reduction of primitive cell size. (**B**) AFE-FE phase transition in PZO under the action of an electric field along the [111] direction from our developed fully first-principles approach. Phonon frequency–dependent WP3 (**C**), phonon-phonon scattering rate (**D**), and group velocity (**E**) for AFE and FE PZO.

emergence of the ferroelectric (FE) component, as indicated by the increase in remnant polarizations (fig. S7A), elevates the thermal conductivity k.

In addition to switching ratio and number of switching cycles, another important factor for high-performance thermal switching is switching time. To evaluate switching time, we applied a trapezoidal wave with rise time of 250 ns. We simultaneously measured the switching current (Fig. 3E), showing a switching time of ~150 ns. Compared to other thermal switching mechanisms such as light-triggered alignment of molecular chains (on the order of tens of seconds) (*13*) and electrochemical controlled ion intercalation (ranging from tens of minutes to hours) (*16*–*18*), the switching time is reduced by more than eight orders of magnitude. This demonstrates the advantages of PZO-based thermal switch devices in the control of phonon transport.

Atomic-scale analysis of thermal switching

To explain electric field–triggered thermal switching in PZO, we performed in situ synchrotron–based 3D-RSM experiments under an electric field (*47*) to characterize the structural evolution during the AFE-FE phase transition. At the Q_z position of PZO (222) under zero field and beyond transition field, the in-plane RSM can be extracted from the 3D data of the AFE and FE phases, respectively (Fig. 4A). The upper panel shows clear 1/4 satellite peaks originating from the ↑↑↓↓ antipolar dipole arrangement, which can substantially multiply the primitive cells. In addition, we observed the splitting of the main Bragg peak PZO (222), which can be attributed to the pseudo-tetragonal features of *Pbam* PZO. When we applied the electric field beyond the transition field, the 1/4 satellite peaks disappeared and the split main Bragg peaks merged into one as the structure transformed from AFE to FE, clearly indicating the notable structure differences between AFE and FE phases as well as the reduction of primitive cell size or the number of atoms in primitive cell. Thus, we report a different mechanism of reversibly modulating primitive cell size to realize high-performance thermal switching.

To further understand the mechanism of thermal switching and explain the film orientation–dependent switching ratio, we employed a full first-principles approach (*47*) to simulate the AFE-FE phase transition. The similar polarization along the AFE-FE phase transition path between theoretical calculations and experimental data (fig. S8) indicates that the phase transition behavior is successfully simulated by our first-principles approach. As shown in Fig. 4B, the electric field along the [111] direction can substantially decrease n, from 40 to 10. More importantly, the ↑↑↓↓ antipolar dipole arrangement disappears, consistent with Fig. 4A. We also simulated electric fields along the [001] and [110] directions and the FE phases of PZO have n larger than 10 (fig. S9). This is why (111)-oriented PZO film has the highest switching ratio compared with other oriented PZO films (see Fig. 3D and fig. S5B).

We then employed dual-phonon theory (*48*) to calculate the thermal conductivity of PZO. It should be noted that our theoretical calculations accounted for the three-phonon process. Phonon renormalization and the four-phonon process were omitted because of their minimal influence on thermal conductivity (fig. S10, supplementary note 4). The calculated thermal conductivity of AFE PZO at room temperature stands at 0.97 W m^{-1} K^{-1}, which is marginally higher than experimental data (0.55 to 0.80 W m^{-1} K^{-1}). Possible reasons might include sample defects, boundary scattering, and substrate effects. Furthermore, we meticulously examined the impact of primitive cell and supercell sizes on thermal conductivity in addition to conducting comprehensive convergence tests (supplementary notes 5 and 6). We compare the weighted phonon-phonon scattering phase space (WP3) for the AFE structure at zero field and FE structure under an electric field applied along the [111] direction (Fig. 4C). Here, WP3 measures the scattering phase space of populated phonons at the target temperature (300 K) (*49*). Notably, as n decreases from 40 to 10, the value of WP3 has a substantial decrease over the entire phonon spectrum. Such decrease of WP3 is different from previous strategies (*8*) for modulating the thermal conductivity, for which phonons within only a narrow spectrum range are modulated. Because thermal conductivity is the integral of phonon transport over the entire phonon spectrum (*8*), the change in WP3 over the entire phonon spectrum is the underlying reason for the much higher switching ratio in the current work than that in previous literature (*21*–*23*, *32*, *33*). In addition to WP3-related phonon lifetime or phonon-phonon scattering rate (Fig. 4D), thermal conductivity is also determined by specific heat, phonon group velocity (Fig. 4E), and phonon anharmonicity (*8*) (fig. S11 and supplementary note 2). Detailed first-principles calculations (supplementary

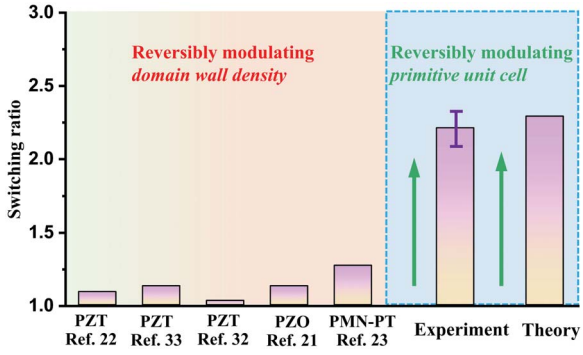


Fig. 5. Thermal switching ratios in the current work and previous FE materials at room temperature. The switching ratios in the experiment are from in situ TDTR measurements and in theory are from the prediction of first-principles calculations with dual-phonon theory (fig. S12), i.e., normal and diffuson-like phonons. In contrast to the previous approach of reversible modulation of domain wall density, our strategy involving the reversible modulation of the number of atoms (n) within the primitive unit cell demonstrates a significantly enhanced switching ratio in antiferroelectric/ferroelectric materials. Experimental switching ratio is representative of at least three independent experiments, each with similar results. Bar, mean ± standard error.

note 3) show that compared with n-dependent WP3, these three other factors all have a much weaker contribution to the thermal conductivity difference between the AFE and FE structures (table S1). Overall, the high switching ratio in PZO arises from the wide-range modulation of the number of atoms n in the primitive unit cell.

We compare our thermal switching ratios with previous AFE/FE materials (*21*–*23*, *32*, *33*) at room temperature (Fig. 5). The fundamental reason behind our high ratios lies in the utilization of the reversible modulation of the number of atoms (n) within the primitive unit cell. This means that materials with n having a large difference before and after the phase transition such as some lead-based perovskites that have incommensurate structures (*50*) may be good candidates for thermal switching. In the previous work, domain wall density was modulated to achieve thermal switching. However, domain sizes in FE materials are typically tens to hundreds of nanometers, much longer than the phonon mean free path at room temperature. For example, as shown in fig. S13, phonons with mean free path less than 14 nm (0.74 nm) contribute to 80% of the total thermal conductivity of PZO from the linearized Boltzmann transport equation (dual-phonon theory) at room temperature. In such cases, modulating domain size or domain wall density can only slightly change the thermal conductivity because Umklapp phonon scattering is much stronger than phonon-domain wall scattering. For our observations, the dominate Umklapp phonon scattering was directly modulated during the electric field–triggered AFE-FE phase transition, resulting in the high switching ratio.

Conclusions

In summary, we report a low-voltage–driven high-contrast, fast-speed, and long-lifetime thermal switching in antiferroelectric PZO by employing a novel mechanism of reversibly modulating the number of atoms (n) within the primitive cell. The thermal switching was realized by simply turning on or off the external electric field, which means there are no moving parts. This would help facilitate the integration of PZO–based thermal switches with other systems. These findings advance the understanding of phonon transport in (anti-)ferroelectrics and provides an efficient strategy to realize active control of heat transfer.

REFERENCES AND NOTES

1. Z. H. Zheng *et al.*, *Nat. Sustain.* **6**, 180–191 (2023).
2. S. Pandya *et al.*, *Nat. Mater.* **17**, 432–438 (2018).
3. L. E. Bell, *Science* **321**, 1457–1461 (2008).
4. Y. Li *et al.*, *Nat. Rev. Mater.* **6**, 488–507 (2021).
5. G. Wehmeyer, T. Yabuki, C. Monachon, J. Q. Wu, C. Dames, *Appl. Phys. Rev.* **4**, 041304 (2017).
6. M. Terraneo, M. Peyrard, G. Casati, *Phys. Rev. Lett.* **88**, 094302 (2002).
7. D. G. Cahill *et al.*, *Appl. Phys. Rev.* **1**, 011305 (2014).
8. X. Qian, J. Zhou, G. Chen, *Nat. Mater.* **20**, 1188–1202 (2021).
9. W. Luo *et al.*, *Phys. Rev. B* **107**, L241107 (2023).
10. T. Du *et al.*, *Nat. Commun.* **12**, 4915 (2021).
11. C. Cazorla, M. Stengel, J. Íñiguez, R. Rurali, *NPJ Comput. Mater.* **9**, 97 (2023).
12. X. Meng *et al.*, *Phys. Rev. Lett.* **122**, 155901 (2019).
13. J. Shin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5973–5978 (2019).
14. N. Varela-Domínguez *et al.*, *J. Mater. Chem. C* **11**, 4588–4594 (2023).
15. Q. Lu *et al.*, *Nat. Mater.* **19**, 655–662 (2020).
16. A. Sood *et al.*, *Nat. Commun.* **9**, 4510 (2018).
17. G. Zhu *et al.*, *Nat. Commun.* **7**, 13211 (2016).
18. J. Cho *et al.*, *Nat. Commun.* **5**, 4035 (2014).
19. S. Lee *et al.*, *Science* **355**, 371–374 (2017).
20. R. Shrestha *et al.*, *Sci. Adv.* **5**, eaax3777 (2019).
21. K. Aryana *et al.*, *Nat. Commun.* **13**, 1573 (2022).
22. J. F. Ihlefeld *et al.*, *Nano Lett.* **15**, 1791–1795 (2015).
23. A. Negi *et al.*, *Adv. Mater.* **35**, e2211286 (2023).
24. Y. Jiang *et al.*, *Nat. Mater.* **21**, 779–785 (2022).
25. P. Sharma *et al.*, *Sci. Adv.* **3**, e1700512 (2017).
26. L. J. McGilly, P. Yudin, L. Feigl, A. K. Tagantsev, N. Setter, *Nat. Nanotechnol.* **10**, 145–150 (2015).
27. J. R. Whyte *et al.*, *Adv. Mater.* **26**, 293–298 (2014).
28. R. G. P. McQuaid, M. P. Campbell, R. W. Whatmore, A. Kumar, J. M. Gregg, *Nat. Commun.* **8**, 15105 (2017).
29. M. A. Weilert, M. E. Msall, A. C. Anderson, J. P. Wolfe, *Phys. Rev. Lett.* **71**, 735–738 (1993).
30. E. Langenberg *et al.*, *Nano Lett.* **19**, 7901–7907 (2019).
31. J. A. Seijas-Bellido *et al.*, *Phys. Rev. B* **96**, 140101 (2017).
32. B. L. Wooten *et al.*, *Sci. Adv.* **9**, eadd7194 (2023).
33. B. M. Foley *et al.*, *ACS Appl. Mater. Interfaces* **10**, 25493–25501 (2018).
34. C. H. Liu, Y. F. Chen, C. Dames, *Phys. Rev. Appl.* **11**, 044002 (2019).
35. A. Negi *et al.*, *Adv. Sci.* **10**, e2301273 (2023).
36. C. H. Liu, P. Lu, Z. Z. Gu, J. K. Yang, Y. F. Chen, *J. Phys. Chem. C* **124**, 26144–26152 (2020).
37. J. A. Seijas-Bellido, H. Aramberri, J. Íñiguez, R. Rurali, *Phys. Rev. B* **97**, 184306 (2018).
38. A. K. Tagantsev *et al.*, *Nat. Commun.* **4**, 2229 (2013).
39. C. A. Randall, Z. Fan, I. Reaney, L.-Q. Chen, S. Trolier-McKinstry, *J. Am. Ceram. Soc.* **104**, 3775–3810 (2021).
40. Y. Yao *et al.*, *Adv. Mater.* **35**, e2206541 (2023).
41. R.-J. Jiang *et al.*, *Nano Lett.* **23**, 1522–1529 (2023).
42. A. R. Chaudhuri *et al.*, *Phys. Rev. B.* **84**, 054112 (2011).
43. P. Dufour *et al.*, *Appl. Phys. Rev.* **10**, 021405 (2023).
44. R. Xu *et al.*, *Adv. Mater.* **35**, e2210562 (2023).
45. S. Lisenkov, Y. L. Yao, N. Bassiri-Gharb, I. Ponomareva, *Phys. Rev. B* **102**, 104101 (2020).
46. A. Fernandez *et al.*, *Adv. Mater.* **34**, e2108841 (2022).
47. Materials and methods are available as supplementary materials.
48. Y. Luo, X. Yang, T. Feng, J. Wang, X. Ruan, *Nat. Commun.* **11**, 2554 (2020).
49. W. Li, N. Mingo, *Phys. Rev. B.* **91**, 144304 (2015).
50. D. Rusu *et al.*, *Nature* **602**, 240–244 (2022).

ACKNOWLEDGMENTS

The authors thank the Scientific Computing Center of Nanjing Normal University and the Big Data Center of Southeast University for performing the numerical calculations presented in this paper. The authors also thank X. Tian for plotting Fig. 1A and Y. Zhao for analyzing the thermal measurement data. **Funding:** This project was finally supported by National Natural Science Foundation of China (grants 52206092, 52035003, and 52372105), Natural Science Foundation of Jiangsu Province (grant BK20210565), Department of Science and Technology of Jiangsu Province (BK20220032), Guangdong Basic and Applied Basic Research Foundation (Grant No. 2020B1515020029), Shenzhen Science and Technology Innovation project (Grant No. JCYJ20200109112829287), and Shenzhen Science and Technology Program (Grant No. KQTD20200820113045083). C.H.L. was finally funded by the Basic Science (Natural Science) Research Project of Higher Education Institutions of Jiangsu Province (21KJB470009), Nanjing Science and Technology Innovation Project for Overseas Students, "Shuangchuang" Doctor program of Jiangsu Province (JSSCBS20210315), and open research fund of Jiangsu Key Laboratory for Design and Manufacture of Micro-Nano Biomedical Instruments, Southeast University (KF202010). Z.H.C. has been supported by "the Fundamental Research Funds for the Central Universities" (Grant No. HIT.OCEF.2022038) and "Talent Recruitment Project of Guangdong" (Grant No. 2019QN01C202). L.B. thanks the Vannevar Bush Faculty Fellowship (VBFF) Grant No. N00014-20-1-2834 from the Department of Defense". Z.L.L. was supported by the National Natural Science Foundation of China (Grant No. 12175235). S.D. acknowledges Science and Engineering Research Board (SRG/2022/000058) and Indian Institute of Science start up grant for financial support. The authors thank the staff from Shanghai Synchrotron Radiation Facility (SSRF) at BL02U2. **Author contributions:** Z.H.C. and C.H.L. conceived and directed the project. C.H.L. performed the theoretical calculations with the assistance of N.B.F., B.X., and L.B. Y.Y.S. prepared the samples, performed the conventional XRD and electrical measurements with the assistance of Y.J.L. Synchrotron measurements were performed and analyzed by Y.Q.D., Y.Y.S., and Z.L.L. H.Z. did the thermal measurements with the assistance of W.C. S.Q.D. carried out STEM experiments. C.H.L., Y.Y.S., and Z.H.C. wrote the manuscript. Z.H.C. and Y.F.C. supervised the project. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj9669
Materials and Methods
Supplementary Notes 1 to 7
Figs. S1 to S18
Table S1
References (*51*–*68*)

Submitted 26 July 2023; accepted 8 November 2023
10.1126/science.adj9669

VACCINES

HIV vaccines induce CD8$^+$ T cells with low antigen receptor sensitivity

Stephen A. Migueles[1], Danielle M. Nettere[1]†, Noah V. Gavil[1]‡, Lawrence T. Wang[1], Sushila A. Toulmin[1], Elizabeth P. Kelly[1], Addison J. Ward[1]§, Siying Lin[1], Sarah A. Thompson[1], Bennett A. Peterson[1], Cassidy S. Abdeen[1], Carina R. Sclafani[1], Patrick F. Pryal[1], Benjamin G. Leach[1], Amanda K. Ludwig[1], Daniel C. Rogan[1], Paulina A. Przygonska[1], Angela Cattani[1], Hiromi Imamichi[1], Abraham Sachs[2], Gal Cafri[2], Ning-Na Huang[1], Andy Patamawenu[1], C. Jason Liang[3], Claire W. Hallahan[3], Diane M. Kambach[4], Edward X. Han[5], Tiffany Coupet[5], Jonathan Chen[5], Susan L. Moir[1], Tae-Wook Chun[1], Emily E. Coates[6], Julie Ledgerwood[6], Julien Schmidt[7,8], Marie Taillandier-Coindard[7,8], Justine Michaux[7,8], HuiSong Pak[7,8], Michal Bassani-Sternberg[7,8], Nicole Frahm[9]¶, M. Juliana McElrath[9], Mark Connors[1]*

Current HIV vaccines designed to stimulate CD8$^+$ T cells have failed to induce immunologic control upon infection. The functions of vaccine-induced HIV-specific CD8$^+$ T cells were investigated here in detail. Cytotoxic capacity was significantly lower than in HIV controllers and was not a consequence of low frequency or unaccumulated functional cytotoxic proteins. Low cytotoxic capacity was attributable to impaired degranulation in response to the low antigen levels present on HIV-infected targets. The vaccine-induced T cell receptor (TCR) repertoire was polyclonal and transduction of these TCRs conferred the same reduced functions. These results define a mechanism accounting for poor antiviral activity induced by these vaccines and suggest that an effective CD8$^+$ T cell response may require a vaccination strategy that drives further TCR clonal selection.

Designing a highly effective HIV/AIDS vaccine to provide long-term protective immunity remains a major global health priority. There are reasons to be optimistic that vaccine-generated CD8$^+$ T cells could play an important role in containing HIV below detectable levels. In chronic infection, robust correlates of spontaneous immunologic control have been identified (*1*). Based upon preclinical work in SIV- or SHIV-challenged Rhesus macaques, CD8$^+$ T cells induced by vaccines or passive transfer of antibodies during acute infection appear to fully suppress or even eradicate lentiviral infection (*2*–*4*). Despite these advances, the correlates of this activity are not fully understood.

Emphasis has shifted away from vaccines that primarily induce HIV-specific CD8$^+$ T cells because of a lack of efficacy demonstrated in some high-profile studies. In the Step trial, the Merck replication defective trivalent adenovirus (Ad) serotype 5 *gag/pol/nef* subtype B vaccine did not protect from infection and did not reduce HIV load after infection (*5*, *6*). We observed low cytotoxic capacity against primary autologous HIV-infected CD4$^+$ T cells among vaccinees receiving this regimen (*7*), although the mechanism was not defined.

The functions of HIV-specific CD8$^+$ T cells induced by current or prior candidate HIV vaccines in humans were investigated in detail. In comparison with chronically HIV-infected long-term nonprogressors/elite controllers (LTNP/ECs) and progressors, the HIV-specific CD8$^+$ T cells of vaccinees had T cell receptors (TCRs) of insufficient sensitivity to efficiently degranulate in response to low peptide–MHC (pMHC) levels on the surface of HIV-infected CD4$^+$ T cells. These results define a mechanism accounting for the poor cytotoxic capacity and antiviral efficacy induced by these vaccines and suggest potential immune response targets for HIV-1 vaccines and immunotherapies.

The Merck Ad5/HIV vaccine induces low CD8$^+$ T cell cytotoxic capacity

Peripheral blood mononuclear cells (PBMCs) of 34 Step vaccinees (fig. S1 and table S1) were examined in a blinded manner. CD8$^+$ T cell effectors were stimulated with autologous HIV-infected CD4$^+$ T cells for 6 days to enable cytotoxic granule loading (*8*, *9*). Cytotoxic responses, measured by a fluorescence microscopy-based technique ("Pathway ICE", infected CD4$^+$ T cell elimination) (fig. S2 and movie S1), strongly correlated with flow cytometry ("FACS ICE") results (r = 0.9, $P < 0.001$) (fig. S3A) and were highly reproducible (pooled SD = 45.9) (fig. S3B). Vaccinee CD8$^+$ T cell cytotoxic capacity was extremely low (median 27.9%), significantly less than that of both progressors (median 44.3%, $P < 0.001$) and LTNP/ECs (median 68.3%, $P < 0.001$) (Fig. 1A). This was not due to lower percentages of vaccinee HIV-specific CD8$^+$ T cells measured by flow cytometry, which were similar to progressors (medians 6.73 versus 6.09%, respectively, $P > 0.5$) (Fig. 1B and fig. S3C). This latter result—a median increase of 5.2 times in HIV-specific CD8$^+$ T cell frequency from a low 0.59% baseline (table S1)—suggested that vaccinee HIV-specific CD8$^+$ T cells had significantly expanded. Despite this apparent expansion, the cytotoxic activity per cell of vaccinees was well below progressors and LTNP/ECs ($P < 0.001$) (Fig. 1C), even after the effector-to-target (E:T) ratio was adjusted to maximize overlap (fig. S3D). Cytotoxic capacity in these vaccinees was most significantly correlated with the frequencies of degranulating (CD107a$^+$) CD8$^+$ T cells (r = 0.49, $P = 0.003$) (Fig. 1D and fig. S3, E and F). Although the responses of vaccinees were not directly compared with ART-suppressed progressors, we previously found that virus suppression does not restore cytolytic activity (*10*, *11*). Thus, the Merck Ad5/HIV vaccine induces HIV-specific CD8$^+$ T cells with robust expansion potential but poor cytotoxic capacity.

Next, we evaluated whether vaccinee CD8$^+$ T cell responses were associated with HIV control. Post-infection PBMCs were available from 14 vaccinees with "low" ($n = 6$: <400 to 5743 copies/ml) or "high" ($n = 8$: 39,345 to 256,125 copies/ml) set point HIV-1 RNA levels (fig. S4 and table S1). Cytotoxic capacity increased significantly following infection in both viral load subgroups ($P \leq 0.03$) but remained significantly less than LTNP/ECs ($P \leq 0.01$) (fig. S4A) despite comparable frequencies of HIV-specific CD8$^+$ T cells ($P \geq 0.05$) (fig. S4B). In addition, the cytotoxic capacity per cell post infection of vaccinees remained significantly lower than that of LTNP/ECs ($P \leq 0.01$) (fig. S4, C and D). Although two vaccinees with <400 HIV RNA copies/ml were among the high responders, preinfection and postinfection cytotoxic capacity in vaccinees did not correlate with set point viral loads ($P > 0.5$) (fig. S5, A and B), likely because the vaccine induced very low cytotoxic capacity overall. The frequency of degranulating CD107a$^+$ CD8$^+$ T cells preinfection was the strongest correlate of set point viral load (r = −0.5, $P = 0.003$) (Fig. 1E and fig. S5, C to G).

Low cytotoxic activity of Ad5/HIV vaccinee CD8$^+$ T cells is not due to reduced cytotoxic protein expression or function

We next investigated the expression and function of cytotoxic proteins. PBMCs from 12 Merck

[1]Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. [2]Surgery Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [3]Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. [4]Agilent Technologies, Inc., Santa Clara, CA, USA. [5]IsoPlexis, Inc., Branford, CT, USA. [6]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. [7]Ludwig Institute for Cancer Research, University of Lausanne, Lausanne, Switzerland. [8]Department of Oncology, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland. [9]Vaccine and Infectious Disease Division and the HIV Vaccine Trials Network, Fred Hutchinson Cancer Center, Seattle, WA, USA.

***Corresponding author. Email: mconnors@niaid.nih.gov**

†Present address: Duke University School of Medicine, Durham, NC, USA.

‡Present address: Center for Immunology, University of Minnesota Medical School, Minneapolis, MN, USA.

§Present address: University of Massachusetts TH Chan School of Medicine, Worcester, MA, USA.

¶Present address: Bill & Melinda Gates Medical Research Institute, Cambridge, MA, USA.

Ad5/HIV vaccinees who had undergone leukapheresis in the HVTN 071 study (fig. S1) were stimulated with HIV-infected CD4$^+$ T cell targets and analyzed by flow cytometry. Relatively low cytotoxic capacity in Ad5/HIV vaccinee CD8$^+$ T cells by FACS ICE occurred despite high PRF or GrB content that did not differ from LTNP/EC CD8$^+$ T cells, except for higher PRF mean fluorescence intensity (MFI) in vaccinees (Fig. 2A and fig. S6A). The fraction of degranulating CD8$^+$ T cells in vaccinees again correlated significantly with cytotoxic capacity (r = 0.76, P = 0.007) (Fig. 2B). Extracellular cytotoxic protein concentrations and activities were measured after coincubating CD8$^+$ T cells and HIV-infected CD4$^+$ T cell targets for 1 hour. Median extracellular PRF and GrB concentrations of vaccinees were intermediate between LTNP/ECs (P = 0.01 and P = 0.005, respectively) and progressors (P > 0.05 and P = 0.01, respectively) (Fig. 2, C and D). Functional activities of PRF (*12*) and GrB exhibited similar trends (Fig. 2, E and F). Extracellular PRF and GrB concentrations correlated directly with cytotoxic activities in all participants (r = 0.7, P < 0.001 and r = 0.9, P < 0.001, respectively) (fig. S6, B and C). For HVTN 071 vaccinees, PRF concentration did not correlate with PRF activity (r = 0.08, P > 0.5). This was probably related to the same peak in activity at low concentrations observed in LTNP/ECs. However, GrB concentration and activity were strongly directly correlated (r = 0.84, P = 0.004) (Fig. 2, G and H). Thus, the cytotoxic proteins that were up-regulated and exocytosed from Ad5/HIV vaccinee CD8$^+$ T cells are functionally intact.

Ad5/HIV vaccinee HIV-specific CD8$^+$ T cells are less multifunctional and exhibit reduced PRF secretory capacity relative to LTNP/EC cells

We also performed an exploratory analysis of the secretome and intracellular signalome (*13*). The CD8$^+$ T cell responses of two LTNP/ECs could be readily distinguished from two Ad5/HIV vaccinees in three-dimensional space by t-distributed stochastic neighbor embedding (t-SNE) (fig. S7A). The principal drivers differentiating these responses were multifunctional subsets enriched in LTNP/ECs that included combinations of the effector proteins granzyme B, MIP-1β, PRF, and RANTES (fig. S7B). A panel of signal transduction phosphoproteins did not differ in frequency between LTNP/ECs and Ad5/HIV vaccinees, suggesting comparable activation status and signal transduction capacity between the groups (fig. S7C). Although frequencies of intracellular plus secreted PRF were similar between Ad5/HIV vaccinees and LTNP/ECs (23.8 versus 39.8%, respectively, P > 0.05), the secreted PRF frequency from Ad5/HIV vaccinee cells was significantly lower than in LTNP/ECs (6.4 versus 29.2%, respectively, P = 0.02) (fig. S7D). Thus, the reduced functionality of Ad5/HIV vaccinee cells relative to those of LTNP/ECs may be related to fewer PRF-containing multifunctional subsets and reduced PRF secretory capacity.

Effector CD8$^+$ T cells of vaccinees degranulate poorly in response to HIV-infected CD4$^+$ T cell targets

Low CD107a$^+$ CD8$^+$ T cell frequencies in response to infected targets may be caused by low numbers of antigen-specific cells or impaired effector-cell degranulation. To control for this, we used an HLA class I–HIV-1 tetramer panel labeling HIV-specific CD8$^+$ T cells (fig. S8, tables S2 and S3, and data S1). The median fractions of HVTN 071 vaccinee tetramer$^+$ CD8$^+$ T cells expressing PRF (75.1%) and GrB (85.2%) following 6-day stimulation with HIV-infected targets were lower than those of LTNP/ECs (89.4%, P = 0.006 and 94.1%, P = 0.02, respectively) and higher than progressors

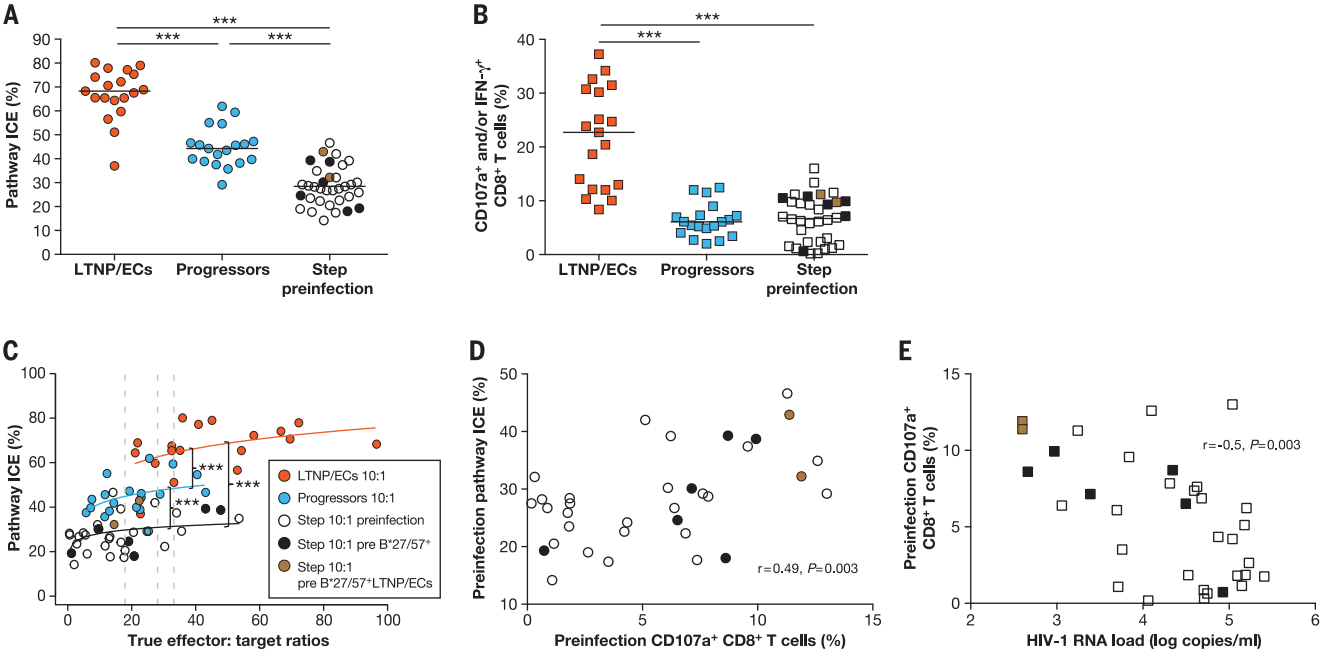


Fig. 1. The Merck Ad5/HIV vaccine induces low cytotoxic capacity per cell but degranulating CD107a$^+$ CD8$^+$ T cell frequencies correlate inversely with HIV-1 load. (**A**) and (**B**) Summary of cytotoxic responses measured by fluorescence microscopy as the elimination of HIV_{GFP}-infected autologous CD4$^+$ T cell targets in 40 min (Pathway ICE, circles) (A) and Day 6 CD107a$^+$ and/or IFN-γ$^+$ CD8$^+$ T cell frequencies measured by flow cytometry (squares) (B) are shown for LTNP/ECs (red symbols, n = 19), progressors (blue symbols, n = 19), B*27$^-$/B*57$^-$ Step vaccinees (open black symbols, n = 26), B*27$^+$ or B*57$^+$ Step vaccinees with HIV-1 RNA levels >400 copies/ml (solid black symbols, n = 6) and B*27$^+$ or B*57$^+$ Step vaccinees with HIV-1 RNA levels ≤400 copies/ml (brown symbols, n = 2). These data are representative of two independent experiments. Horizontal lines indicate median values. Comparisons were made using the Wilcoxon two-sample test. Only significant P values are shown: ***, P ≤ 0.001. (**C**) ICE responses were plotted against the true E:T ratios to assess per-cell cytotoxic capacity. Group differences were quantified by regression analysis. (**D** and **E**) Correlations of CD107a$^+$ CD8$^+$ T cell frequencies with Pathway ICE and HIV-1 RNA levels were determined by the Spearman rank method.

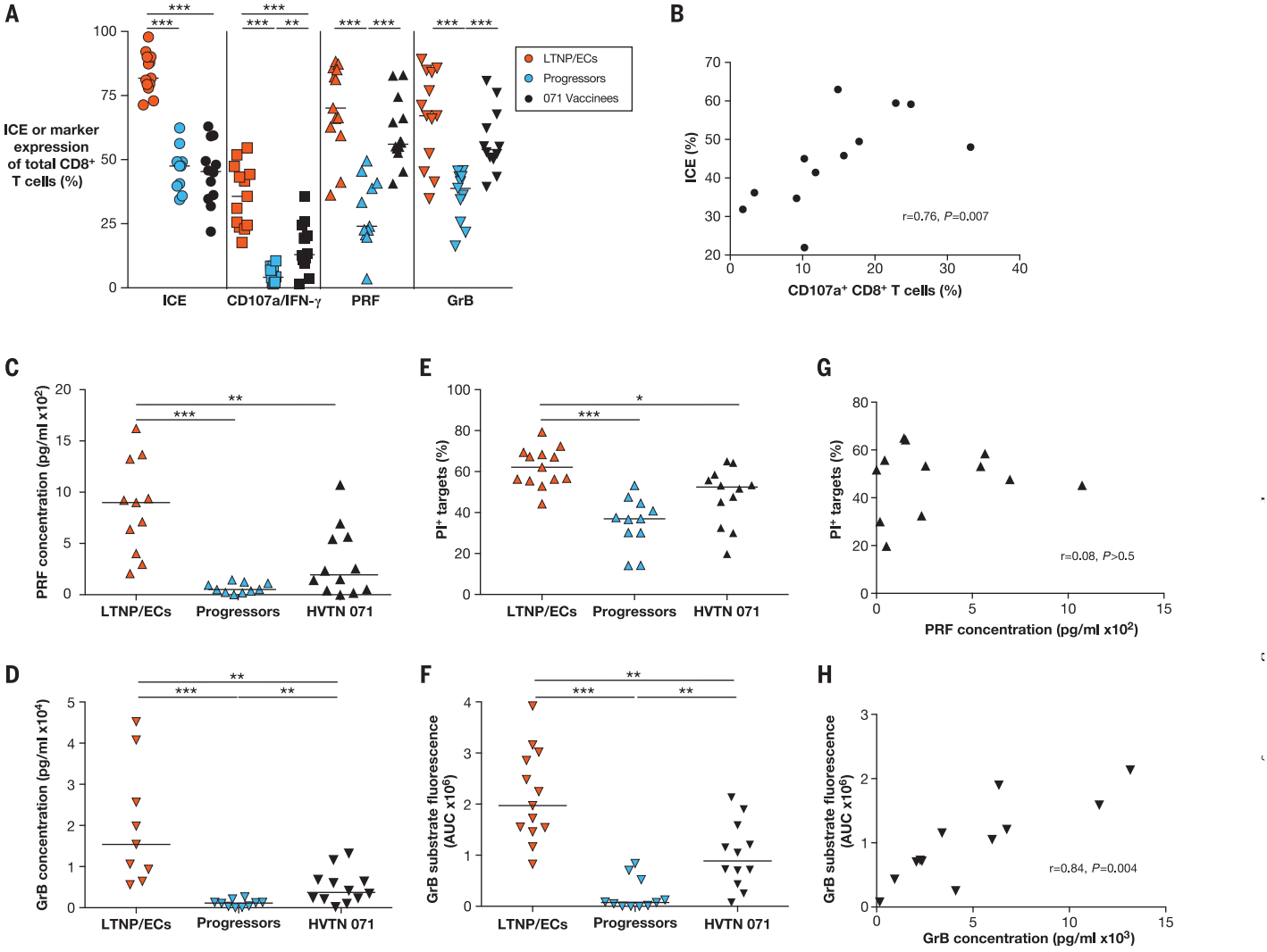


Fig. 2. Low cytotoxic activity of Ad5/HIV vaccinee CD8$^+$ T cells is not due to reduced cytotoxic protein expression or function. (**A**) Flow cytometry was used to measure ICE responses (circles) and day 6 CD8$^+$ T cells expressing CD107a and/or IFN-γ (squares), perforin (PRF, upward triangles) and granzyme B (GrB, downward triangles) from LTNP/ECs (red symbols, *n* = 13), progressors (blue symbols, *n* = 11), and HVTN 071 Ad5/HIV vaccinees (black symbols, *n* = 12). (**B**) In HVTN 071 vaccinees, ICE responses correlated directly with CD107a$^+$ CD8$^+$ T cell frequencies by the Spearman rank method. (**C** and **D**) Concentrations of extracellular PRF (C) and GrB (D) were determined in supernatants by ELISA kits. (**E**) PRF activity in cell pellets was measured by flow cytometry as the percentages of live, propidium iodide (PI)$^+$ CD4$^+$ T cell targets. (**F**) Extracellular GrB activity was measured by changes in fluorescence of a fluorogenic GrB substrate in cell-free supernatants. These data are representative of three independent experiments. Horizontal lines indicate median values. Comparisons were made using the Wilcoxon two-sample test. Only significant *P* values are shown: *, $P \le 0.05$; **, $P \le 0.01$; ***, $P \le 0.001$. (**G** and **H**) In 071 vaccinees, PRF (G) and GrB (H) extracellular concentrations were correlated with their activities by the Spearman rank method.

(59.1%, $P = 0.02$ and 64.5%, $P < 0.001$, respectively) (Fig. 3A). Despite the high cytotoxic protein content of HVTN 071 vaccinee tetramer$^+$ CD8$^+$ T cells, the median fraction expressing the degranulation marker CD107a following 6-hour restimulation with HIV-infected targets was much lower (8.4%) than both progressors (26.2%, $P = 0.04$) and LTNP/ECs (52%, $P < 0.001$) (Fig. 3A). Similar differences for IFN-γ-expressing cells were noted among the groups. We also measured CD8$^+$ T cell degranulation capacity in three additional vaccine cohorts: healthy subjects who had received one dose of the VRC-replication-defective Ad5/HIV vaccine administered as a single IM injection (VRC_006) (*14*) or as a boost after priming with three multi-antigen DNA vaccinations (VRC_008) (*15*) and vaccinees in HVTN 087 who were primed with three multi-antigen DNA vaccinations (ProfectusVax) plus high-dose (1500 μg) IL-12 before boosting with a live, attenuated vesicular stomatitis virus expressing HIV-1 Gag (VSV-Gag) (Fig. 3, fig. S1, and table S2) (*16*). Except for three participants in HVTN 087 with responses that overlapped with LTNP/ECs, HIV-specific cells from vaccinees in these cohorts exhibited very low degranulation capacity (Fig. 3A).

As a second measurement of degranulation, net changes in GrB MFI in effectors upon restimulation (GrB ΔMFI) (fig. S8, upper quadrants of last column) correlated inversely with frequencies of CD107a$^+$ tetramer$^+$ cells ($r = -0.79$, $P < 0.001$) (Fig. 3B). There was markedly reduced loss of GrB in HVTN 071 (+416) and VRC_006/008 (−376) vaccinees compared with chronically infected participants (LTNP/ECs: −13,002; progressors: −4010, $P \le 0.006$) (Fig. 3C). ICE responses confirmed low cytotoxic capacity in the CD8$^+$ T cells of most vaccinees (fig. S9). Considerable homology was observed among sequences corresponding to the synthetic epitopes,

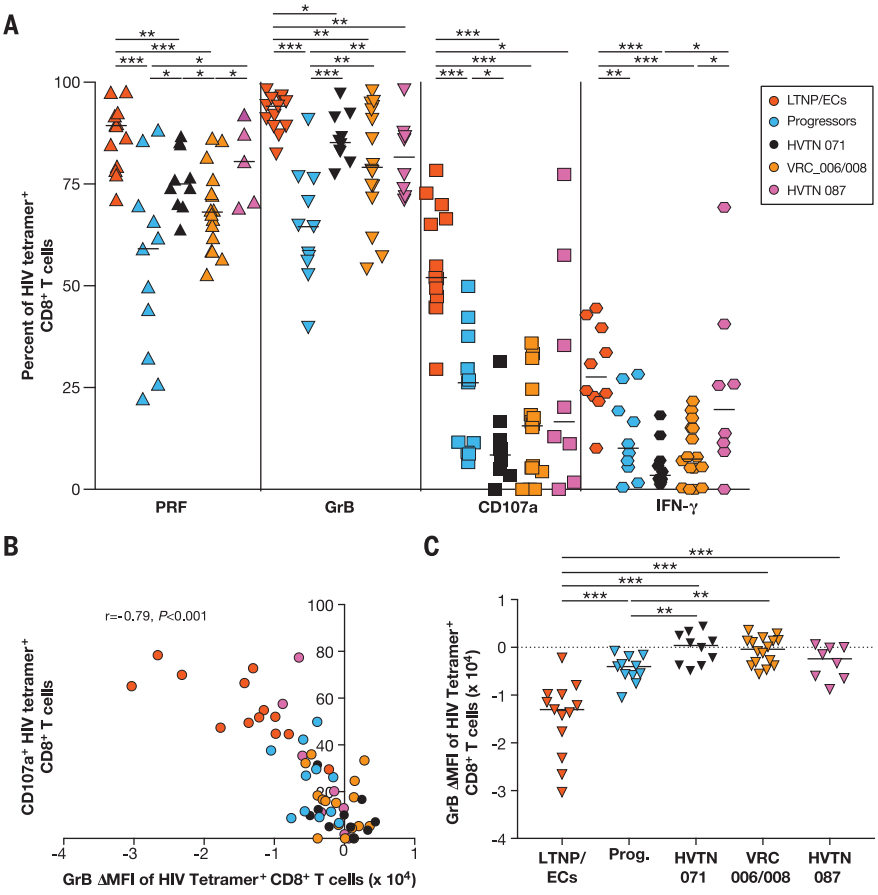


Fig. 3. Effector CD8$^+$ T cells of vaccinees degranulate poorly in response to HIV-infected CD4$^+$ T cells. (**A**) Summary of the percent expression in HIV tetramer$^+$ CD8$^+$ T cells of PRF (upward triangles), GrB (downward triangles), CD107a (squares), and IFN-γ (hexagons) measured by flow cytometry after 6-day stimulation with HIV-infected CD4$^+$ T cell targets are shown in LTNP/ECs (red symbols, n = 10 to 13), progressors (blue symbols, n = 10 to 11), HVTN 071 Ad5/HIV vaccinees (black symbols, n = 10), VRC_006/008 Ad5/HIV vaccinees (orange symbols, n = 16), and HVTN 087 vaccinees (magenta symbols, n = 5 to 8). For CD107a and IFN-γ expression, Day 6 CD8$^+$ T cells were restimulated with autologous uninfected or HIV-infected CD4$^+$ T cell targets for 6 hours in the presence of monensin, brefeldin-A, and a CD107a monoclonal antibody. Background responses to uninfected targets were subtracted. (**B**) In gated HIV tetramer$^+$ CD8$^+$ T cells that were GrB$^+$ and/or CD107a$^+$, loss of GrB MFI upon restimulation (GrB ΔMFI) inversely correlated with CD107a$^+$HIV tetramer$^+$ CD8$^+$ T cells by the Spearman rank method. (**C**) Summary of GrB ΔMFI in HIV tetramer$^+$ CD8$^+$ T cells is shown for the same subjects in (A). These data are representative of two independent experiments. Horizontal lines indicate median values. Comparisons were made using the Wilcoxon two-sample test. Only significant P values are shown: *, $P \le 0.05$; **, $P \le 0.01$; ***, $P \le 0.001$.

the HIV-1 vaccine inserts, HIV_{SF162} (used to superinfect autologous targets), and HXB2 (as a reference), which suggested that poor responsiveness was unlikely, due to lack of HIV_{SF162} recognition (fig. S10). Thus, the HIV-specific CD8$^+$ T cells induced by various HIV vaccines exhibit a markedly impaired ability to degranulate upon restimulation with HIV-infected CD4$^+$ T cells.

Reduced degranulation capacity of Ad5/HIV vaccine-induced HIV-specific CD8$^+$ T cells is due to low avidity and impaired cytotoxic granule polarization toward the immunologic synapse

Cells from LTNP/ECs, progressors, and Ad5/HIV vaccinees were restimulated for 6 hours with a range of single optimal peptide dilutions (corresponding to their immunodominant tetramer$^+$ CD8$^+$ T cell responses) and assessed for CD107a expression (Fig. 4 and fig. S11A). The peptide concentrations that induced CD107a responses in tetramer$^+$ CD8$^+$ T cells halfway between baseline and maximum (EC_{50}) were similar between LTNP/ECs and progressors (medians −8.81 versus −8.68 log M, $P > 0.05$), but more than a log-fold lower in chronically infected participants compared with Ad5/HIV vaccinees (−7.44, $P \le 0.001$) (Fig. 4A). Responses restricted by protective HLA class I proteins enriched in LTNP/ECs tended to be of higher avidity across participant groups (*17*, *18*), although the avidity of other dominant responses were more variable (Fig. 4A). Given the comparable avidity of cells from LTNP/ECs and progressors—which is consistent with prior results (*1*, *19*)—our investigations subsequently focused on features distinguishing the responses of LTNP/ECs from vaccinees. CD107a expression was significantly lower across the range of peptide concentrations in Ad5/HIV vaccinees compared with LTNP/ECs ($P < 0.001$) (Fig. 4B). In pairwise comparisons, differences were most pronounced at lower concentrations (median difference 43.39% at a 10^{-8} dilution, $P < 0.001$ versus 2.73% at a 10^{-5} dilution, $P > 0.05$). Similarly, degranulation responses across all peptide dilutions in CD8$^+$ T cells specific for cytomegalovirus, influenza virus, respiratory syncytial virus and, in COVID-19-mRNA vaccine recipients, SARS-CoV-2 exceeded those of Ad5/HIV vaccinees ($P < 0.001$) (Fig. 4C and table S3).

The peptide concentration present on infected targets was interpolated by plotting CD107a responses in HIV tetramer$^+$ CD8$^+$ T cells measured in parallel with infected targets on peptide response curves in six subjects (fig. S11B). This suggested the effective optimal epitope concentration on infected targets was approximately 10^{-8} M, the peptide concentration at which very few Ad5/HIV vaccinee cells degranulated (median 10.9%, mean 20.6%) (Fig. 4B). An analysis of the peptide concentration in 1 to 1.7×10^7 sorted HIV-infected CD4$^+$ T cells by targeted mass spectrometry showed that two peptides reliably detected by this technique (EI8 and FL8) were below the 10^{-6} M detection limit (fig. S12), consistent with the functional data. Thus, the peptide concentration on the infected cell surface appears to be approximately 200 to 1000 times below that commonly used in most in vitro measures of CD8$^+$ T cells. Furthermore, low antigen sensitivity is most pronounced at these very low interpolated peptide concentrations on HIV-infected cells.

Reduced functional avidity has been associated with altered mobilization of cytotoxic granules toward the immunologic synapse (IS) (*20*–*22*). Using a fluorescent microscopy platform (Fig. 4D), lytic granule polarization to the IS was significantly lower in vaccinee versus LTNP/EC CD8$^+$ T cells at low (10^{-8} M) but not high (10^{-5} M) peptide concentrations ($P = 0.038$) (Fig. 4E). However, the numbers of effector–target conjugates formed in LTNP/ECs and Ad5/HIV vaccinees were similar ($P > 0.05$) (fig. S11C), consistent with prior work (*20*, *21*). Conventional confocal microscopy revealed similar findings, with lower numbers of polarized conjugates in Ad5/HIV vaccinees compared to LTNP/ECs at low peptide concentrations (fig. S13). Thus, low-avidity interactions of vaccinee HIV-specific CD8$^+$ T cells induced limited cytotoxic granule polarization toward the IS.

Low-sensitivity TCRs on CD8$^+$ T cells induced by Ad5/HIV vaccines account for low avidity interactions leading to reduced functionality

Poor functional avidity may be a result of the binding strength of the TCR for its pMHC ligand, TCR density, dependency for the CD8 coreceptor or adhesion molecule ligation, or signaling cascade defects downstream of the TCR complex (*23*–*25*). When PBMCs from LTNP/ECs or Ad5/HIV vaccinees were stimulated with peptide pools, expression on HIV tetramer$^+$ cells of the co-stimulatory receptors CD2, CD8β, CD11a, and CD18 did not differ ($P > 0.5$) (fig. S14A), nor did degranulation responses to a broad range of plate-bound CD3-specific antibody concentrations ($P > 0.5$) (Fig. 5A). These results, paired with earlier phosphoprotein data (fig. S7C), suggested that costimulatory molecule expression or diminished signal transduction were unlikely explanations for low PRF polarization and degranulation in Ad5/HIV vaccinee cells.

Therefore, it remained possible that a feature of the TCR complex itself could underpin the reduced functional avidity observed in Ad5/HIV vaccinees. TCR sequence analysis (*26*) of immunodominant HIV tetramer$^+$ CD8$^+$ T cells revealed significantly greater diversity in the TCR repertoires of Ad5/HIV vaccinees compared with LTNP/ECs (0.83 versus 0.51 Simpson's Diversity Index, respectively, $P = 0.02$) (Fig. 5, B and C, and table S4). Based only on TCRβ sequences, a few shared and/or public clonotypes in responses restricted by the protective class I proteins B27 or B57 were observed in both groups (table S4). TCRs were then transduced into Jurkat cells or primary PBMCs to investigate function, focusing on dominant clonotypes (*27*). Peptide–MHC multimers with reversible Ni^{2+}-nitrilotriacetic acid histidine tags (NTAmers) (*28*) revealed a nonsignificant trend toward more rapid dissociation of B27-KK10–specific monomers from the TCRs of three Ad5/HIV vaccinees versus more delayed off rates of TCRs from two LTNP/ECs (median K_{off}, 13.6 versus 40.1, respectively, $P > 0.05$) (fig. S14B). To further examine this with additional TCR clones, the degranulation capacity of primary PBMCs from a healthy donor transduced with single TCRs was measured in response to heterologous HIV-infected targets matched only at the presenting HLA class I protein. Dominant Ad5/HIV vaccinee TCRs on tetramer$^+$ effectors exhibited significantly less CD107a up-regulation in response to HIV-infected targets than did LTNP/EC TCR-bearing effectors (32.1% versus 50.3%, $P = 0.01$) (Fig. 5D), despite similar tetramer staining ($P > 0.5$, fig. S14C). Consistent with the earlier avidity results, B27/KK10-specific TCRs were characterized by greater degranulation responses irrespective of participant group. However, dominant LTNP/EC TCRs recognizing antigens presented by even nonprotective HLA proteins mediated robust degranulation responses. Thus, low-avidity interactions of HIV-specific CD8$^+$ T cells in Ad5/HIV vaccinees result from a preponderance of less sensitive TCRs, which induce limited cytotoxic granule polarization toward the IS, inefficient degranulation, and reduced cytotoxic capacity in response to the low peptide concentrations presented by HIV-infected CD4$^+$ T cell targets.

Discussion

We observed that even after 6 day stimulation to enable lytic granule loading (*8*, *9*), cytolytic killing induced by several past or present candidate vaccines was universally low relative to chronically infected participants as a result of low TCR antigen sensitivity that was insufficient to respond to the low levels of pMHC on HIV-infected primary CD4$^+$ T cells. High-avidity cells have been associated with greater

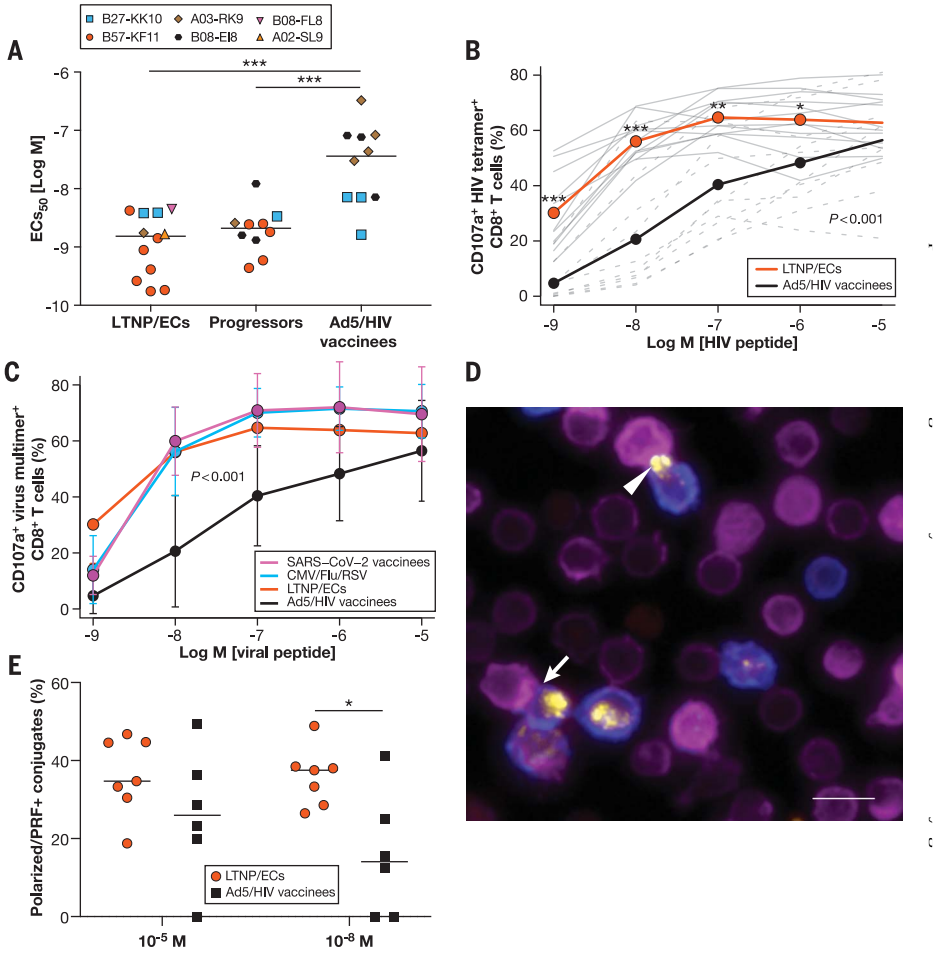


Fig. 4. Reduced degranulation of Ad5/HIV vaccinee CD8$^+$ T cells is due to low functional avidity and restricted cytotoxic granule polarization. (**A**) Half-maximal effective concentrations (EC_{50}) of peptides inducing CD107a expression in HIV tetramer$^+$ CD8$^+$ T cells from LTNP/ECs ($n = 12$), progressors ($n = 10$), and Ad5/HIV vaccinees ($n = 10$) were estimated by a nonlinear model for six immunodominant responses: B27-KK10 (blue symbols), B57-KF11 (red symbols), A03-RK9 (brown symbols), B08-EI8 (black symbols), B08-FL8 (magenta symbols), and A02-SL9 (orange symbols). (**B**) Responses were compared across all dilutions for LTNP/ECs (median red line) and Ad5/HIV vaccinees (median black line) by the generalized estimating equations (GEE) method. Responses at each peptide dilution were compared using the Wilcoxon two-sample test. (**C**) LTNP/EC and vaccinee responses were compared by the GEE method with the following ones specific for CMV ($n = 4$), Flu ($n = 4$), RSV ($n = 2$), and SARS-CoV-2 in mRNA vaccine recipients ($n = 7$): A02-CMV NV9 + A02-Flu GL9 + B07-RSV NL9 (median blue solid line) and A02-SARS-CoV-2 YL9 (median solid magenta line). These data are representative of two independent experiments. (**D**) A representative micrograph shows F-actin$^+$ (magenta) cells forming conjugates between CD8$^-$ (CD4$^+$) and CD8$^+$ (blue) T cells containing bright PRF (yellow) in a polarized (arrowhead) or nonpolarized (arrow) pattern. Scale bar 10 μm. (**E**) Frequencies of polarized conjugates for LTNP/ECs (red symbols, $n = 7$) and vaccinees (black symbols, $n = 6$) in response to high (10^{-5}) or low (10^{-8} M) epitope concentrations were compared by the Wilcoxon two-sample test. Data were generated from a single experiment. Horizontal lines indicate median values. Only significant P values are designated: *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$.

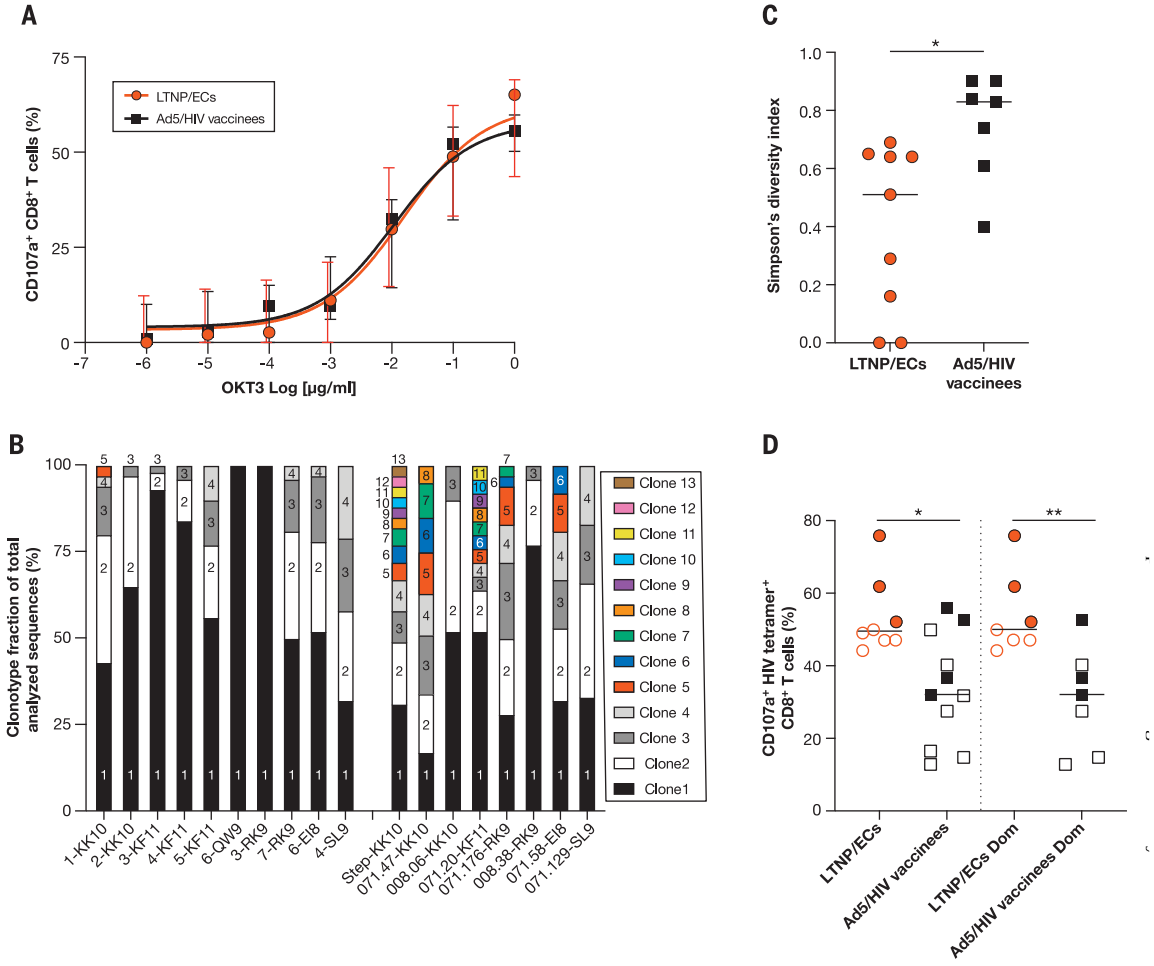


Fig. 5. Low-sensitivity TCRs on CD8$^+$ T cells induced by Ad5/HIV vaccines account for low-avidity interactions. (**A**) Peptide-expanded CD8$^+$ T cells from LTNP/ECs (red line, n = 6) and Ad5/HIV vaccinees (black line, n = 6) were restimulated with plate-bound anti-CD3 at multiple dilutions, assessed for CD107a expression in HIV tetramer$^+$ cells and compared by the GEE method. (**B**) TCR clonal composition of HIV tetramer$^+$ CD8$^+$ T cells from 7 LTNP/ECs (10 specificities) and 8 vaccinees (8 specificities) shown as total number (range, 1 to 13) and fraction of each TCR clonotype in the epitope-specific repertoire, abbreviated as follows: B27-KK10 (KK10), B57-KF11 (KF11), B57-QW9 (QW9), A03-RK9 (RK9), B08-EI8 (EI8), and A02-SL9 (SL9). (**C**) After standardizing for sampling differences, clonotypic diversity was calculated by Simpson's diversity index and compared between LTNP/ECs (red circles, n = 9) and vaccinees (black squares, n = 7). (**D**) HIV-negative donor PBMCs transduced with single TCRs from LTNP/ECs [red circles, n = 8 total, seven dominant ("Dom") clones] or vaccinees (black squares, n = 11 total, seven Dom clones) were analyzed for CD107a expression in HIV tetramer$^+$ cells by flow cytometry. Solid symbols designate B27-restricted, KK10-specific TCR clones. Background responses to uninfected targets were subtracted. These data are representative of two independent experiments. Horizontal lines indicate median values. Unless otherwise noted, comparisons were made using the Wilcoxon two-sample test. *, $P \le 0.05$; **, $P \le 0.01$.

antigen clearance compared to low-avidity cells (*21*, *24*, *29*–*31*). Further investigation will be needed to define the precise nature of the functions mediated by these low-sensitivity TCRs at the molecular level, which may relate to biophysical properties such as the response to mechanical force loaded on to the TCR–pMHC bond (*32*).

Although high-affinity TCRs are thought to be rapidly selected after acute infection (*33*), the situation is likely to be different for HIV. Maturation of the CD8$^+$ T cell response has been observed following vaccination of experimental animals or humans (*34*–*39*) and during the interruption of treatment (*19*). We observed increases in cytotoxic capacity in some vaccinees after HIV infection. The few vaccinees who developed the highest degranulation and cytotoxic responses in our study were recipients of a replication-competent, attenuated, recombinant VSV vaccine. Although not examined, it is possible that induction of low-avidity CD8$^+$ T cells contributed to the recent failure in efficacy trials of a replication-defective Ad26 vector (NCT03060629). Thus, the polyclonal, low-avidity responses we observed may benefit from further clonal selection by additional vaccinations, persisting vectors, and/or vectors resulting in low pMHC ligands to preferentially select high-avidity responses.

Identification of the measures of CD8$^+$ T cell function that reliably correlate with vaccine efficacy is a high priority for the development of HIV vaccines that stimulate a CD8$^+$ T cell response. CD8$^+$ T cell proliferation and cytotoxic protein accumulation can be added to T cell frequencies and IL-2, IFN-γ, and TNF-α polyfunctionality as candidates unlikely to accurately predict vaccine-induced immune control (*7*, *40*, *41*). The 2- to 10-μM peptide concentration used in most assays is in 200 to 1000-fold excess of the 1 to 10 nM that would mimic the quantity of pMHC complexes on the infected cell surface. Although measures of cytokine-producing cells have been useful for enumerating HIV-specific T cells, our data suggest that assays which provide a more physiologic CD8$^+$ T cell stimulus and measure functions more closely tied to antiviral efficacy should be adapted for use at the scale of clinical vaccine trials.

Although some HIV vaccine trials have eroded interest in pursuing T cell–based vaccines, they should not be viewed as absolute indicators of CD8$^+$ T cell antiviral capacity. Rather, it is possible that newer, more immunogenic vaccines and measurements that more closely correlate with antiviral efficacy may guide approaches that can more fully demonstrate the antiviral efficacy of CD8$^+$ T cells in humans.

REFERENCES AND NOTES

1. S. A. Migueles, M. Connors, *Nat. Immunol.* **16**, 563–570 (2015).
2. N. L. Haigwood *et al.*, *Immunol. Lett.* **51**, 107–114 (1996).
3. S. G. Hansen *et al.*, *Nature* **473**, 523–527 (2011).

4. Y. Nishimura *et al*., *Nature* **543**, 559–563 (2017).
5. S. P. Buchbinder *et al*., *Lancet* **372**, 1881–1893 (2008).
6. D. W. Fitzgerald *et al*., *J. Infect. Dis.* **203**, 765–772 (2011).
7. S. A. Migueles *et al*., *PLOS Pathog.* **7**, e1002002 (2011).
8. S. A. Migueles *et al*., *Nat. Immunol.* **3**, 1061–1068 (2002).
9. S. A. Migueles *et al*., *J. Virol.* **94**, e01595-20 (2020).
10. S. A. Migueles *et al*., *Immunity* **29**, 1009–1021 (2008).
11. S. A. Migueles *et al*., *J. Virol.* **83**, 11876–11889 (2009).
12. D. Keefe *et al*., *Immunity* **23**, 249–262 (2005).
13. J. Rossi *et al*., *Blood* **132**, 804–814 (2018).
14. A. T. Catanzaro *et al*., *J. Infect. Dis.* **194**, 1638–1649 (2006).
15. B. S. Graham *et al*., *PLOS ONE* **8**, e59340 (2013).
16. S. S. Li *et al*., *Clin. Vaccine Immunol.* **24**, 11 (2017).
17. J. R. Almeida *et al*., *Blood* **113**, 6351–6360 (2009).
18. C. T. Berger *et al*., *J. Virol.* **85**, 9334–9345 (2011).
19. S. Viganò *et al*., *PLOS Pathog.* **9**, e1003423 (2013).
20. M. Koneru, D. Schaer, N. Monu, A. Ayala, A. B. Frey, *J. Immunol.* **174**, 1830–1840 (2005).
21. M. R. Jenkins, A. Tsun, J. C. Stinchcombe, G. M. Griffiths, *Immunity* **31**, 621–631 (2009).
22. A. M. Beal *et al*., *Immunity* **31**, 632–642 (2009).
23. S. Viganò *et al*., *Clin. Dev. Immunol.* **2012**, 153863 (2012).
24. M. Nauerth *et al*., *Sci. Transl. Med.* **5**, 192ra87 (2013).
25. M. Hebeisen *et al*., *Front. Immunol.* **6**, 582 (2015).
26. R. Yossef *et al*., *JCI Insight* **3**, e122467 (2018).
27. J. A. Conrad *et al*., *J. Immunol.* **186**, 6871–6885 (2011).
28. J. Schmidt *et al*., *J. Biol. Chem.* **286**, 41723–41735 (2011).
29. M. A. Alexander-Miller, G. R. Leggatt, J. A. Berzofsky, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4102–4107 (1996).
30. T. Lövgren *et al*., *Cancer Immunol. Immunother.* **61**, 817–826 (2012).
31. M. S. Abdel-Hakeem, M. Boisvert, J. Bruneau, H. Soudeyns, N. H. Shoukry, *PLOS Pathog.* **13**, e1006191 (2017).
32. E. L. Reinherz, W. Hwang, M. J. Lang, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2215694120 (2023).
33. M. K. Slifka, J. L. Whitton, *Nat. Immunol.* **2**, 711–717 (2001).
34. R. P. Johnson, *Vaccine* **20**, 1985–1987 (2002).
35. Z. Hu *et al*., *J. Virol.* **88**, 5356–5368 (2014).
36. M. J. Estcourt *et al*., *Int. Immunol.* **14**, 31–37 (2002).
37. C. Ranasinghe *et al*., *J. Immunol.* **178**, 2370–2379 (2007).
38. D. K. Wijesundara *et al*., *PLOS ONE* **9**, e105366 (2014).
39. S. Adnan *et al*., *PLOS Pathog.* **11**, e1004633 (2015).
40. M. J. McElrath *et al*., *Lancet* **372**, 1894–1905 (2008).
41. H. Perrin, G. Canderan, R. P. Sékaly, L. Trautmann, *Curr. Opin. HIV AIDS* **5**, 368–376 (2010).
42. S. Migueles, HIV Vaccines Induce CD8+ T Cells with Low Antigen Receptor Sensitivity, Version 1, Dryad (2023); https://doi.org/10.5061/dryad.v41ns1s2b.

ACKNOWLEDGMENTS

HLA class I / HIV, CMV, and SARS-CoV-2 tetramers were obtained through the NIH Tetramer Core Facility. We thank S. Rosenberg for sharing resources and technical expertise in the performance and analyses of the TCR experiments. We also thank M. C. Sneller for sharing samples from COVID-19 mRNA vaccine recipients and A. Brown for sharing the full-length *nef*-restored molecular clone pSF162R3 and providing technical advice. We also thank A. Clayton (HVTN), A. Poole (NIH), and C. Rehm (NIH) for their invaluable assistance in procuring the requested biospecimens. **Funding:** This research was supported in part by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases. The HVTN Laboratory Program (NIH U01 AI068618), the Seattle Vaccine Unit (NIH U01 AI069481), and HVTN Core Leadership (U01 AI068614) provided additional support. **Author contributions:** Conceptualization: S.A.M., N.F., M.J.M., and M.C. Methodology: S.A.M., D.M.N., N.V.G., L.T.W., S.A.T., E.P.K., A.J.W., S.L., S.A.T., B.A.P., C.S.A., C.R.S., P.F.P., B.G.L., A.K.L., A.C., H.I., A.S., G.C., N.H., C.J.L., C.W.H., D.K., T.C., J.C., S.L.M., T.-W.C., E.E.C., J.S., M.T.-C., J.M., H.P., and M.B.-S. Investigation: S.A.M., D.M.N., N.V.G., L.T.W., S.A.T., E.P.K., A.J.W., S.L., S.A.T., B.A.P., C.S.A., C.R.S., P.F.P., B.G.L., A.K.L., D.C.R., P.A.P., A.C., A.S., G.C., N.H., A.P., D.M.K., T.C., E.E.C., J.S., M.T.-C., J.M., and H.P. Funding acquisition: J.L., M.J.M., and M.C. Project administration: A.P., J.L., N.F., M.J.M., and M.C. Supervision: S.A.M., D.M.K., T.C., and M.C. Writing – original draft: S.A.M. and M.C. Writing – review and editing: S.A.M., D.M.N., A.S., G.C., C.J.L., E.X.H., J.L., J.S., M.B.-S., N.F., M.J.M., and M.C. **Competing interests:** E.X.H., T.C., and J.C. are employed by and have equity ownership in IsoPlexis, Inc. M.J.M. is an inventor on patent application 63127975 submitted by Fred Hutchinson Cancer Center, Scripps, and IAVI that covers Immunogenic compositions related to immunodominant peptides LumSyn. J.C. is an inventor on patent 10,274,486 held by Yale University, which covers high-throughput multiplexed detection. All other authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The full dataset underlying the results is accessible at Dryad (*42*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg0514
Materials and Methods
Figs. S1 to S14
Tables S1 to S5
References (*43*–*58*)
Movie S1
Data S1

Submitted 5 December 2022; accepted 3 November 2023
10.1126/science.adg0514

VIRAL PALEOGENOMICS

Ancient chicken remains reveal the origins of virulence in Marek's disease virus

Steven R. Fiddaman[1]*†, Evangelos A. Dimopoulos[2,3]†, Ophélie Lebrasseur[4,5], Louis du Plessis[6,7], Bram Vrancken[8,9], Sophy Charlton[2,10], Ashleigh F. Haruda[2], Kristina Tabbada[2], Patrik G. Flammer[1], Stefan Dascalu[1], Nemanja Marković[11], Hannah Li[12], Gabrielle Franklin[13], Robert Symmons[14], Henriette Baron[15], László Daróczi-Szabó[16], Dilyara N. Shaymuratova[17], Igor V. Askeyev[17], Olivier Putelat[18], Maria Sana[19], Hossein Davoudi[20], Homa Fathi[20], Amir Saed Mucheshi[21], Ali Akbar Vahdati[22], Liangren Zhang[23], Alison Foster[24], Naomi Sykes[25], Gabrielle Cass Baumberg[2], Jelena Bulatović[26], Arthur O. Askeyev[17], Oleg V. Askeyev[17], Marjan Mashkour[20,27], Oliver G. Pybus[1,28], Venugopal Nair[1,29], Greger Larson[2]‡, Adrian L. Smith[1]*‡, Laurent A. F. Frantz[30,31]*‡

The pronounced growth in livestock populations since the 1950s has altered the epidemiological and evolutionary trajectory of their associated pathogens. For example, Marek's disease virus (MDV), which causes lymphoid tumors in chickens, has experienced a marked increase in virulence over the past century. Today, MDV infections kill >90% of unvaccinated birds, and controlling it costs more than US$1 billion annually. By sequencing MDV genomes derived from archeological chickens, we demonstrate that it has been circulating for at least 1000 years. We functionally tested the *Meq* oncogene, one of 49 viral genes positively selected in modern strains, demonstrating that ancient MDV was likely incapable of driving tumor formation. Our results demonstrate the power of ancient DNA approaches to trace the molecular basis of virulence in economically relevant pathogens.

Marek's disease virus (MDV) is a highly contagious alphaherpesvirus that causes a tumor-associated disease in poultry. At the time of its initial description in 1907, Marek's disease was a relatively mild disease with low mortality, characterized by nerve pathology that affected mainly older individuals (*1*). However, over the course of the 20th century, MDV-related mortality has risen to >90% in unvaccinated chickens. To prevent this high mortality rate, the poultry industry spends more than US$1 billion per year on health intervention measures, including vaccination (*2*).

The increase in virulence and clinical pathology of MDV infection has likely been driven by a combination of factors. First, the growth in the global chicken population since the 1950s led to more viral replication, which increased the supply of new mutations in the population. In addition, the use of imperfect (also known as "leaky") vaccines that prevent symptomatic disease but do not prevent transmission of the virus likely shifted selective pressures and led to an accelerated rate of MDV virulence evolution (*3*). Combined, these factors have altered the evolutionary trajectory, resulting in modern hyperpathogenic strains. To date, the earliest sequenced MDV genomes were sampled in the 1960s (*4*), several decades after the first reports of MDV causing tumors (*5*). As a result, the genetic changes that contributed to the increas virulence of MDV infection before the 19 remain unknown.

MDV has been circulating in Europe for at least 1000 years

To empirically track the evolutionary change in MDV virulence through time, we generated MDV genome sequences (serotype 1) isolated from the skeletal remains of archeological chickens. We first shotgun sequenced 995 archeological chicken samples excavated from more than 140 western Eurasian archeological sites and screened for MDV reads using HAYSTAC (*6*) with a herpesvirus-specific database. Samples with any evidence of MDV reads were then enriched for viral DNA by using a hybridization-based capture approach based on RNA baits designed to tile the entire MDV genome (excluding one copy of each of the terminal repeats and regions of low complexity). To validate the approach, we also captured and sequenced DNA from the feather of a modern Silkie chicken that presented MDV symptoms. As a negative control, we also included an ancient sample that displayed no evidence of MDV reads after screening (OL1214; Serbia, 14th to 15th century).

Using the capture protocol, we identified 15 ancient chickens with MDV-specific reads of ≥25 base pairs (bp) in length. This approach also yielded a ~4× genome from a modern positive control. We found that the majority (88 to 99%) of uniquely mapped reads that were generated from ancient samples classified as MDV-positive were ≥25 bp, whereas the majority (53 to 100%) of uniquely mapped reads that were generated from samples considered MDV-negative were <25 bp. In addition, samples considered MDV-positive yielded between 308 and 133,885 uniquely mapped reads (≥25 bp), whereas samples considered MDV-negative (including a negative control) (table S2) yielded between 0 and 211 uniquely mapped reads of ≥25 bp. MDV-positive ancient samples ranged in depth of coverage from 0.13× to 41.92× (OL1385) (Fig. 1A and table S2), with seven genomes at ≥2× coverage.

In all positive samples, the proportion of duplicated reads approached 100%, which indicated that almost all of the molecules in each library were sequenced at least once (fig. S1). Reads obtained from MDV-positive ancient samples had chemical signatures of DNA damage typically associated with ancient DNA (fig. S2). By contrast, reads obtained from our modern positive control did not show any evidence of DNA damage (fig. S2). The earliest unequivocally MDV-positive sample (with 4760 postcapture reads ≥25 bp) was derived from a 10th- to 12th-century chicken from eastern France (Andlau in Fig. 1A and table S2). Together, these results demonstrate that MDV strains have been circulating in western Eurasian poultry for at least 1000 years.

[1]Department of Biology, University of Oxford, Oxford, UK. [2]The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK. [3]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. [4]Centre d'Anthropobiologie et de Génomique de Toulouse, CNRS/Université Toulouse III Paul Sabatier, Toulouse, France. [5]Instituto Nacional de Antropología y Pensamiento Latinoamericano, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina. [6]Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. [7]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [8]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium. [9]Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Brussels, Belgium. [10]BioArCh, Department of Archaeology, University of York, York, UK. [11]Institute of Archaeology, Belgrade, Serbia. [12]Institute of Immunity and Transplantation, University College London, London, UK. [13]Silkie Club of Great Britain, Charing, UK. [14]Fishbourne Roman Palace, Fishbourne, UK. [15]Leibniz-Zentrum für Archäologie, Mainz, Germany. [16]Medieval Department, Budapest History Museum, Budapest, Hungary. [17]Laboratory of Biomonitoring, The Institute of Problems in Ecology and Mineral Wealth, Tatarstan Academy of Sciences, Kazan, Russia. [18]Archéologie Alsace–PAIR, Sélestat, Bas-Rhin, France. [19]Departament de Prehistòria, Universitat Autònoma de Barcelona, Barcelona, Spain. [20]Bioarchaeology Laboratory, Central Laboratory, University of Tehran, Tehran, Iran. [21]Department of Art and Architecture, Payame Noor University (PNU), Tehran, Iran. [22]Iranian Ministry of Cultural Heritage, Tourism, and Handicrafts, North Khorasan Office, Iran. [23]Department of Archaeology, School of History, Nanjing University, China. [24]Headland Archaeology, Edinburgh, UK. [25]Department of Archaeology, University of Exeter, Exeter, UK. [26]Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden. [27]CNRS, National Museum Natural History Paris, Paris, France. [28]Department of Pathobiology and Population Sciences, Royal Veterinary College, London, UK. [29]Viral Oncogenesis Group, Pirbright Institute, Woking, UK. [30]Palaeogenomics Group, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilians-Universitat, Munich, Germany. [31]School of Biological and Chemical Sciences, Queen Mary University of London, London, UK.

***Corresponding author. Email: steven.fiddaman@biology.ox.ac.uk (S.R.F.); adrian.smith@biology.ox.ac.uk (A.L.S.); laurent.frantz@lmu.de (L.A.F.F.)**

†These authors contributed equally to this work.

‡These authors contributed equally to this work.

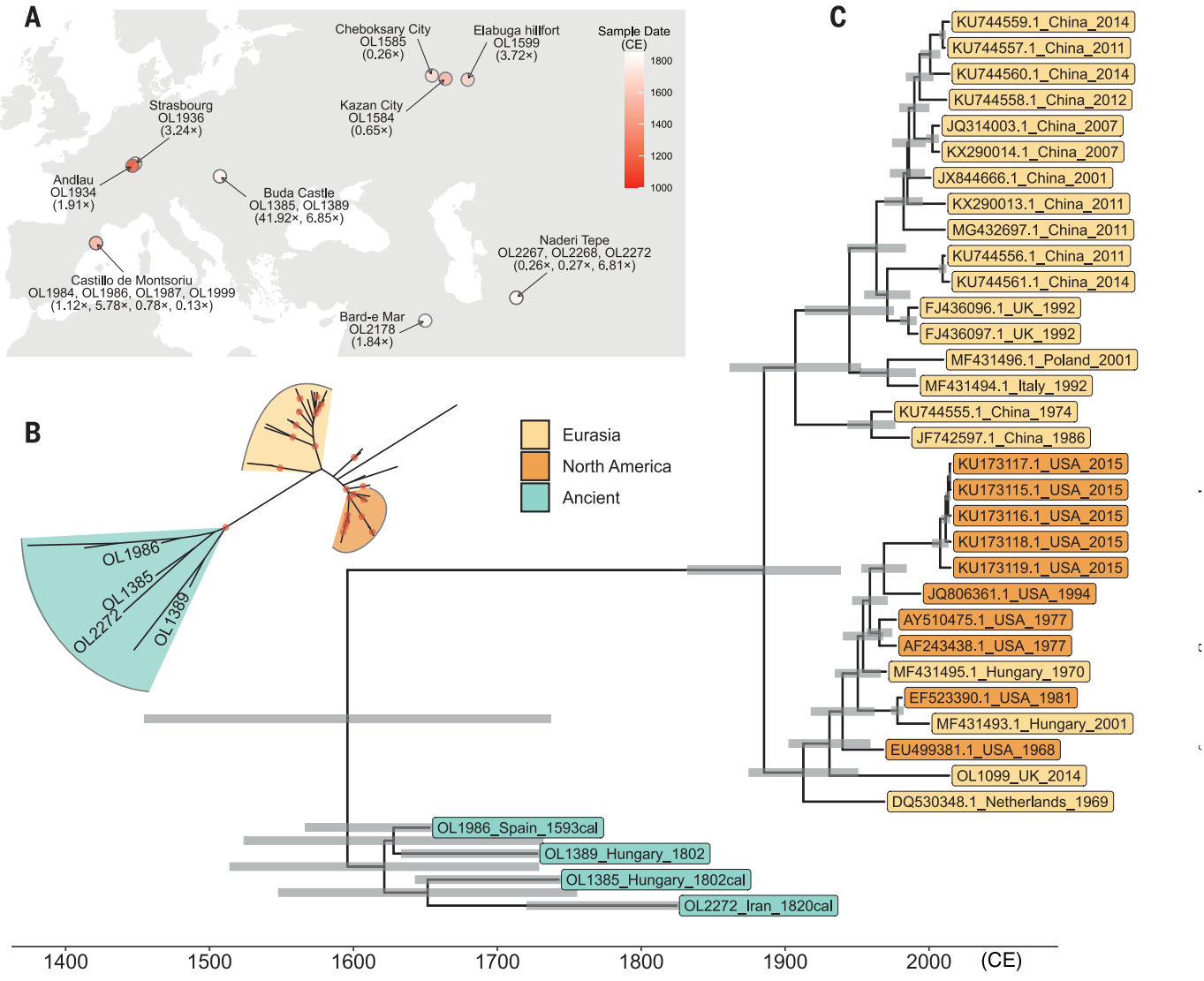


Fig. 1. Locations of MDV-positive samples and time-scaled phylogeny. (**A**) Map showing the locations of screened archeological chicken samples that were positive for MDV sequence. Colored circles indicate sample dates (either from calibrated radiocarbon dating or estimated from archeological context) (table S1). Average sequencing depth after capture is given in parentheses under sample names. If more than one sample was derived from the same site, this is indicated by a list of sample identifiers (beginning "OL") and sequencing depths in parentheses. (**B**) Unrooted NJ tree of 42 modern and 10 ancient genomes. Only the four high-coverage ancient samples used in our BEAST analysis were labeled in this tree (table S2). Nodes with bootstrap support of >90 are indicated with red dots. (**C**) Timescaled maximum clade credibility tree of ancient and modern MDV sequences by using the uncorrelated lognormally distributed (UCLD) relaxed clock and the general time-reversible (GTR) substitution model. Gray bars indicate the 95% HPD for the age of each node. The "cal" suffix for ancient samples indicates that samples were radiocarbon dated and that these date distributions were used as priors for the molecular clock analyses (*24*).

Ancient MDV strains are basal to modern lineages

To investigate the relationship between ancient and modern MDV strains, we built phylogenetic trees based on both neighbor-joining (NJ) and maximum-likelihood (ML) methods. We first built trees using 10 ancient genomes with at least 1% coverage at a depth of ≥5×, a modern positive control derived from the present study (OL1099), and 42 modern genomes from public sources (table S3). Both NJ (Fig. 1B and fig. S3) and ML trees (fig. S4) match the previously described general topology (*7*), in which Eurasian and North American lineages were evident, along with a well-supported (bootstrap 94) ancient clade (Fig. 1B). The same topology was also obtained when we restricted our ML analysis to include only transversion sites (fig. S5). Last, we built a tree using an outgroup (*Meleagrid herpesvirus 1*, accession NC_002641.1) to root our topology (fig. S6). We obtained a well-supported topology showing that the ancient MDV sequences form a highly supported clade that lies basal to all modern MDV strains (including the modern positive control OL1099).

Next, we built a time-calibrated phylogeny using BEAST [v. 1.10 (*8*)] that included 31 modern genomes collected since 1968 (table S3) and four ancient samples with an average depth of coverage >5× (OL1986, Castillo de Montsoriu, Spain, 1593 calibrated CE; OL1385, Buda Castle, Hungary, 1802 calibrated CE; OL1389, an additional Buda Castle sample from the same archeological context as OL1385; and OL2272, Naderi Tepe, Iran, 1820 calibrated CE) (Fig. 1A and tables S1 and S2). The time of the most recent common ancestor (TMRCA) of the phylogeny was 1602 CE [95% highest posterior

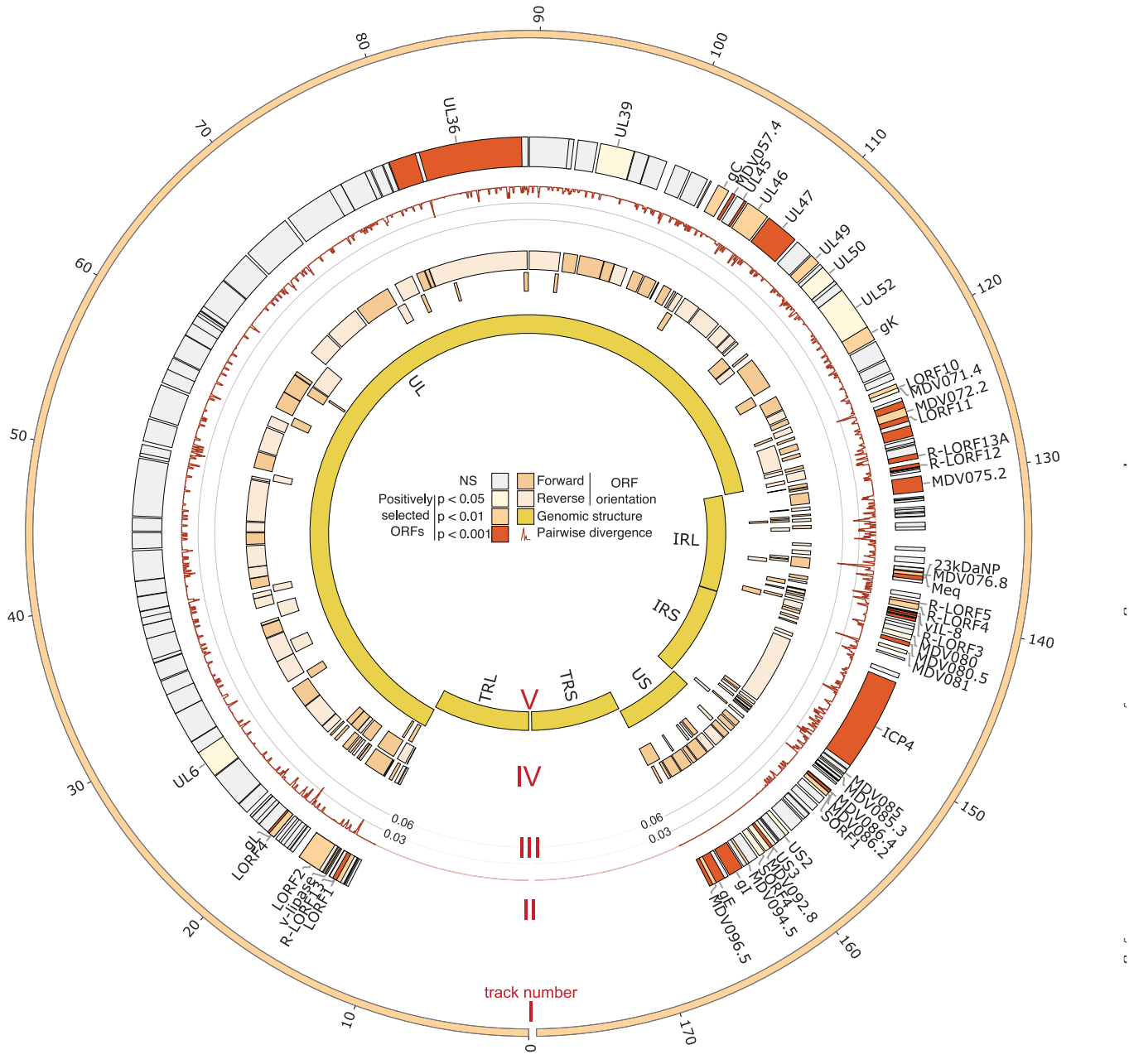


Fig. 2. Branch-site selection analysis of MDV genomes. The MDV genome is represented as a circular structure with gross genomic architecture displayed on the innermost track (track V) and genomic coordinates shown on the outermost track (units, $\times 10^3$ kb; track I). Because the long terminal repeat (TRL) and short terminal repeat (TRS) are copies of the long internal repeat (IRL) and the short internal repeat (IRS), respectively, selection analysis excluded the TRL and the TRS regions, leaving only the unique long (UL) and unique short (US) regions along with the two internal repeats. Results of the positive selection analysis are displayed on track II, where ORFs are shaded according to the strength of statistical support (corrected *P* values) for positive selection. Sliding window average pairwise divergence between ancient and modern samples is shown on track III, and ORF orientation is shown on track IV. NS, not significant.

density (HPD) interval, 1486 to 1767 CE] (Fig. 1C and table S4).

As previously reported (*7*), we found that aside from a few exceptions, most Eurasian and North American MDV strains formed distinct clades (Fig. 1B), which suggests that there has been little recent transatlantic exchange of the virus. The inclusion of time-stamped ancient MDV sequences improved the accuracy of the molecular clock analysis and pushed back the TMRCA of all modern MDV sequences, from 1922–1952 (*7*) to 1881 (95% HPD interval, 1822 to 1929) (table S4). Our mean TMRCA of modern MDV is concordant with a recent estimate that incorporated 26 modern MDV genomes from East Asian chickens [1880; 95% HPD, 1772 to 1968 (*9*)]. This phylogenetic analysis implies that the two major modern clades of MDV were likely established before the earliest documented increases in MDV virulence in the 1920s. Furthermore, because birds infected with highly virulent MDV would not have survived a transatlantic crossing, a TMRCA of 1938 (95% HPD, 1914 to 1958) for the clade containing the earliest North American sample (CU2, 1968; accession EU499381.1) could be consistent with the virus having been transmitted

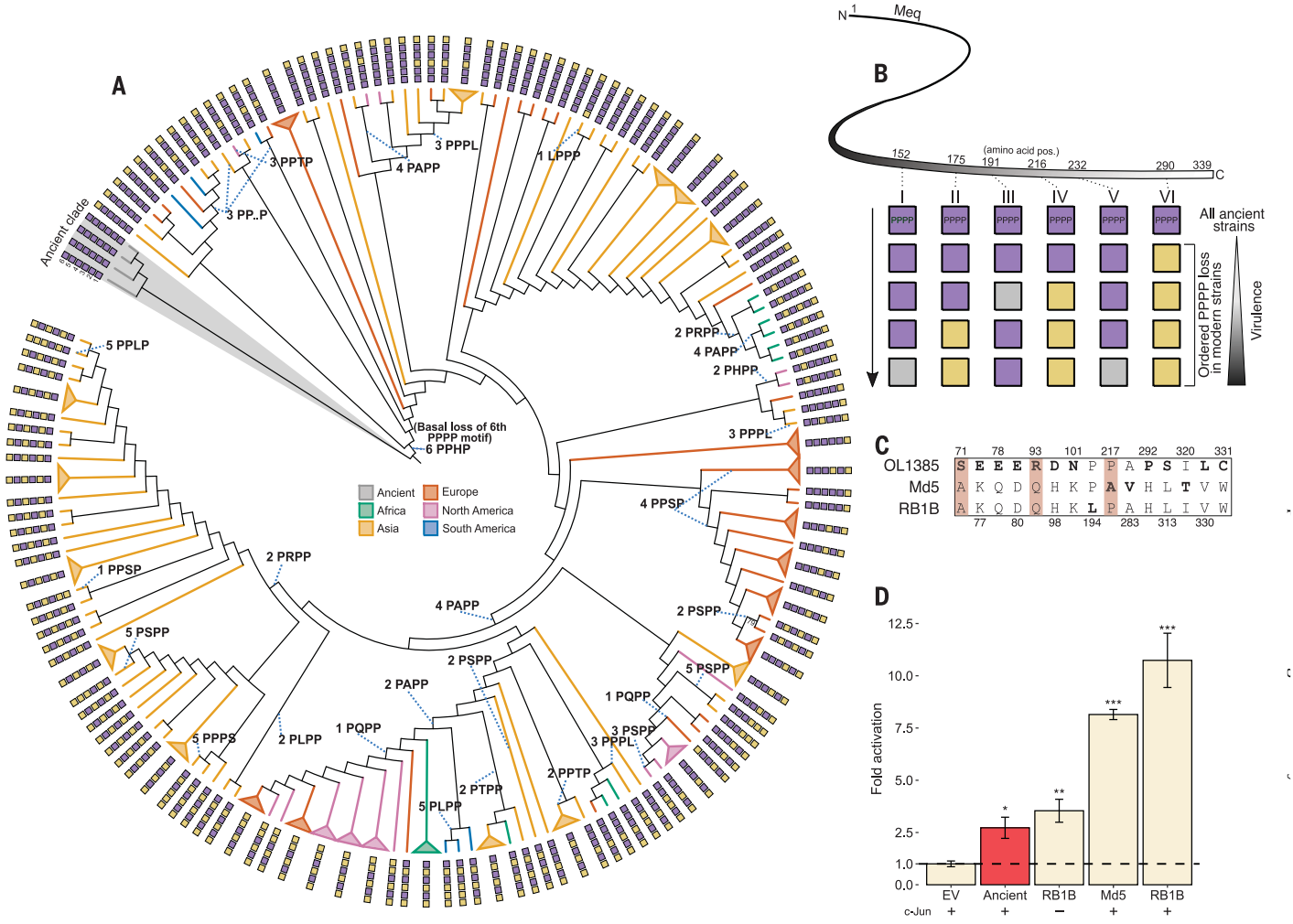


Fig. 3. *Meq* has undergone ordered loss of tetraproline repeats and increased transactivation ability. (**A**) Phylogenetic analysis of 412 *Meq* sequences of standard length (1017 bp). The outermost track shows the integrity of each tetraproline motif (purple squares, intact; yellow squares, disrupted). The mutations that disrupt the tetraproline motif are linked by dotted blue lines (for example, "4 PAPP" indicates that the fourth tetraproline motif is disrupted by a proline-to-alanine substitution in the second proline position; "3 PP..P" denotes a deletion of the third proline in the third tetraproline motif). A complete version of this figure is provided in fig. S7. A, alanine; H, histidine; L, leucine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine. (**B**) Proposed model for the most common ordered loss of tetraproline motifs in *Meq*. Purple and yellow boxes indicate presence and absence of an intact tetraproline, respectively. The gray box on the third row indicates that the third tetraproline is occasionally lost after the sixth, but typically only in terminal branches. The two gray boxes in the bottom row indicate that either the first or fifth tetraproline is lost at this point. (**C**) Positions of amino acid differences between the ancient Hungarian MDV strain (OL1385) and the two modern strains (RB1B and Md5). Positions that were also found to be under positive selection are highlighted in red. (**D**) The transactivation ability of *Meq* reconstructed from an ancient Hungarian MDV strain (OL1385) was compared with the transactivation abilities of modern strains: RB1B and Md5 (very virulent pathotypes). To show the effect of the partner protein c-Jun on transactivation ability, the strongest transactivator, RB1B, was tested with (+) and without (−) c-Jun. Transactivation ability is expressed as fold activation relative to baseline signal from an empty vector (EV). Error bars are standard deviation, and statistical significance was determined by using Dunnett's test for comparing several treatment groups with a control. *P < 0.05; **P < 0.01; ***P < 0.001.

before the most substantial virulence increases leading up to the 1960s. These results are also consistent with the hypothesis that Eurasian and North American MDV lineages independently evolved toward increased virulence (*7*).

Virulence factors are among positively selected genes in the modern MDV lineage

The rapid increase in MDV virulence could potentially have been driven by gene loss or gain, which would have substantially altered the biology of the virus (*10*, *11*). Analysis of a Hungarian, high-coverage, MDV genome (OL1385; >41×) from the 18th to 19th century indicated that it had the full complement of genes present in modern sequences. This indicates that there was no gene gain or loss in either ancient or modern lineage (Fig. 2). We also found that all MDV microRNAs, some of which are implicated in pathogenesis and oncogenesis in modern strains (*12*), were intact and highly conserved in ancient strains (table S5). Together, these results indicate that the acquisition of virulence most likely resulted not from changes in MDV genome content or organization but from point mutations.

Considering sites at which we had coverage for at least two ancient genomes, we identified 158 fixed single-nucleotide polymorphisms between the ancient and modern samples, of which 31 were found in intergenic regions and may be candidates for future study of MDV regulatory regions (table S6). To assess the impact of positive selection on point mutations,

we performed a branch-site analysis in PAML (*13*) (ancient sequences as background lineage, modern sequences as foreground lineage) on open reading frames (ORFs) using four ancient MDV genomes (OL1385, OL1389, OL1986, and OL2272). After we controlled the false discovery rate using the Benjamini-Hochberg procedure (*14*), this analysis identified 49 ORFs with significant evidence for positive selection (Fig. 2 and table S7).

Several positively selected loci identified in this analysis have previously been associated with MDV virulence in modern strains. Some of these are known immune modulators or potential targets of a protective response. This includes ICP4 (infected cell protein 4), a large transcriptional regulatory protein involved in innate immune interference. ICP4 appears to be an important target of T cell–mediated immunity against MDV in chickens that have the B21 major histocompatibility complex (MHC) haplotype (*15*), and it is plausible that sequence variation in important ICP4 epitopes could confer differential susceptibility to infection.

We also identified signatures of positive selection in several genes that encode viral glycoproteins (gC, gE, gI, gK, and gL). Glycoproteins are important targets for the immune response to MDV (*16*). Most MDV peptides presented on chicken MHC class II are derived from just four proteins (*17*), of which two were glycoproteins found to be under selection in our analysis (gE and gI). This result indicates that glycoproteins are likely under selection in MDV because they are immune targets. The limited scope of immunologically relevant MDV peptides presented by MHC class II may have important implications for vaccine development.

Positive selection was also detected in the viral chemokine termed viral interleukin-8 (vIL-8) [considered a functional ortholog of chicken CXC ligand 13 (*18*)]. vIL-8 is an important virulence factor that recruits B cells for lytic replication and $CD4^+$ $CD25^+$ T cells that are transformed to generate lymphoid tumors. Viruses that lack vIL-8 are severely impaired in the establishment of infection and generation of tumors through bird-to-bird transmission (*19*), so sequence variation in this gene could plausibly affect transmission.

The key oncogene of MDV has experienced positive selection and an ordered loss of tetraproline motifs

Our selection scan also identified Meq, a transcription factor considered to be the master regulator of tumor formation in MDV (*20*). The *Meq* coding sequence had the greatest average pairwise divergence between ancient and modern strains across the entirety of the MDV genome (Fig. 2), implying that there were numerous sequence changes along the branch that leads to modern samples. Animal experiments have demonstrated that *Meq* is essential for tumor formation (*20*), and polymorphisms in this gene, even in the absence of variants elsewhere in the genome, are known to confer substantial differences in strain virulence or vaccine breakthrough ability (*21*).

Meq exerts transcriptional control on downstream gene targets (in both the host and the viral genome) through its C-terminal transactivation domain. This domain is characterized by PPPP (tetraproline) repeats spaced throughout the second half of the protein, and the number of tetraproline repeats is inversely proportional to the virulence of the MDV strain (*22*). The difference in the number of tetraproline repeats in most strains is the result of point mutations rather than deletion or duplication; these strains are considered "standard-length" Meq (339 amino acids). In some strains, however, tetraproline repeats have been duplicated ("long" Meq strains, 399 amino acids) or deleted ("short" Meq strains, 298 amino acids; or "very short" Meq, 247 amino acids). These mutations have led to varying numbers of tetraproline repeats between strains.

We did not find any evidence of duplication or deletion in ancient Meq sequences, which indicates that these are standard-length Meq. We then identified point mutations in a database that contains four ancient Meq sequences (OL1385, OL1389, OL1986, and OL2272) along with 408 modern standard-length Meq sequences (table S8). This analysis demonstrated that ancient Meq had six intact tetraproline motifs, whereas all modern standard-length Meq sequences had between two and five. All ancient Meq sequences had a distinctive additional intact tetraproline motif at amino acids 290 to 293. This tetraproline motif was disrupted by a point mutation—causing a change from proline to histidine—in the recent evolutionary history of standard length–Meq MDV strains.

To further explore the virulence-related disruption of tetraprolines in modern *Meq* sequences, we constructed a phylogeny of *Meq* sequences (Fig. 3A). Mapping the tetraproline content of each sequence on the phylogeny indicated that tetraprolines have been lost in a specific order. After the universal disruption of the sixth tetraproline through a point mutation (at amino acids 290 to 293) at the base of the modern MDV lineage, the fourth tetraproline was disrupted at the base of two major lineages (amino acids 216 to 219). Disruption of the fourth tetraproline was followed in seven independent lineages by the disruption of the second tetraproline (amino acids 175 to 178) and then by the loss of either the first (amino acids 152 to 155) or the fifth tetraproline (amino acids 232 to 235) in six lineages (Fig. 3, A and B).

Our analysis indicated that the second and fourth tetraprolines (codons 176 and 217) were under positive selection (table S7). Although there were some observations of virus lineages that exhibited an alternative loss order [for example, the occasional loss of the third tetraproline (amino acids 191 to 194) after the loss of the fourth], such lineages are not widespread, which suggests that they may become stuck in local fitness peaks and are outcompeted by lineages following the order described here. The independent recapitulation of this pattern in different lineages suggests that loss of tetraproline motifs acts as a ratchet, in which each subsequent loss results in an increase in virulence, and once lost, motifs are unlikely to be regained.

Ancient Meq is a weak transactivator that likely did not drive tumor formation

The initial description of Marek's disease in 1907 did not mention tumors (*1*). Given the degree of sequence differentiation observed between ancient and modern *Meq* genes, ancient MDV genotypes could have been incapable of driving lymphoid cell transformation. To test this hypothesis experimentally, we assessed whether ancient Meq had lower transactivation capabilities, compared with modern strains, in a cultured cell–based assay.

To do so, we synthesized an ancient *Meq* gene on the basis of our highest-coverage ancient sample (OL1385; Buda Castle, Hungary; 1802 calibrated CE) and experimentally tested its transactivation function. We also cloned "very virulent" modern pathotype strains (RB1B and Md5), which each differ from ancient Meq at 13 to 14 amino acid positions (Fig. 3C and table S9). All the Meq proteins were expressed in cells alongside a chicken protein (c-Jun), with which Meq forms a heterodimer, and a luciferase reporter containing the Meq binding (AP-1) sequence.

Relative to the baseline signal, the transactivation of the very virulent Meq strains RB1B and Md5 were 7.5 and 10 times greater, respectively (Fig. 3D). Consistent with previous reports (*23*), removal of the partner protein, c-Jun, from RB1B resulted in severe abrogation of the transactivation capability (Fig. 3D). Ancient Meq exhibited a ~2.5-fold increase in transactivation relative to the baseline but was substantially lower (67 to 75%) than that of Meq from the two very virulent pathotypes (Fig. 3D). The ancient Meq was thus a demonstrably weaker transactivator than Meq from modern strains of MDV.

Given that the transcriptional regulation of target genes (both host and virus) by *Meq* is directly related to oncogenicity (*20*, *23*), the weaker transactivation we demonstrate is likely associated with reduced or absent tumor formation. These data indicate that ancient MDV strains were unlikely to cause tumors and were less pathogenic than modern strains. Ancient MDV likely established a chronic infection

characterized by slower viral replication, low levels of viral shedding, and low clinical pathology, which facilitated maximal lifetime viral transmission in preindustrialized, low-density settings.

Conclusions

Overall, our results demonstrate that MDV has been circulating in western Eurasia for at least the past millennium. By reconstructing and functionally assessing ancient and modern genomes, we showed that ancient MDV strains were likely substantially less virulent than modern strains and that the increase in virulence took place over the past century. Along with changes in several known virulence factors, we identified sequence changes in the *Meq* gene—the master regulator of oncogenesis—that drove its enhanced ability to transactivate its target genes and drive tumor formation. The historical perspective that our results provide can form the basis on which to rationally improve modern vaccines and track or even predict future virulence changes. Last, our results highlight the utility of functional paleogenomics to generate insights into the evolution and fundamental biological workings of pathogen virulence.

REFERENCES AND NOTES

1. J. Marek, *Dtsch. Tierarztl. Wochenschr.* **15**, 417–421 (1907).
2. C. Morrow, F. Fehler, in *Marek's Disease*, F. Davison, V. Nair, Eds. (Academic Press, 2004), pp. 49–61.
3. A. F. Read *et al*., *PLOS Biol.* **13**, e1002198 (2015).
4. C. S. Eidson, K. W. Washburn, S. C. Schmittle, *Poult. Sci.* **47**, 1646–1648 (1968).
5. N. Osterrieder, J. P. Kamil, D. Schumacher, B. K. Tischer, S. Trapp, *Nat. Rev. Microbiol.* **4**, 283–294 (2006).
6. E. A. Dimopoulos *et al*., *PLOS Comput. Biol.* **18**, e1010493 (2022).
7. J. Trimpert *et al*., *Evol. Appl.* **10**, 1091–1101 (2017).
8. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
9. K. Li *et al*., *Front. Microbiol.* **13**, 1046832 (2022).
10. K. Majander *et al*., *Curr. Biol.* **30**, 3788–3803.e10 (2020).
11. B. Mühlemann *et al*., *Science* **369**, eaaw8977 (2020).
12. M. Teng *et al*., *J. Gen. Virol.* **96**, 637–649 (2015).
13. Z. Yang, *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
14. Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc.* **57**, 289–300 (1995).
15. A. R. Omar, K. A. Schat, *Virology* **222**, 87–99 (1996).
16. C. J. Markowski-Grimsrud, K. A. Schat, *Vet. Immunol. Immunopathol.* **90**, 133–144 (2002).
17. S. Halabi *et al*., *PLOS Biol.* **19**, e3001057 (2021).
18. S. Haertle *et al*., *Front. Microbiol.* **8**, 2543 (2017).
19. A. T. Engel, R. K. Selvaraj, J. P. Kamil, N. Osterrieder, B. B. Kaufer, *J. Virol.* **86**, 8536–8545 (2012).
20. B. Lupiani *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11815–11820 (2004).
21. A. M. Conradie *et al*., *PLOS Pathog.* **16**, e1009104 (2020).
22. K. G. Renz *et al*., *Avian Pathol.* **41**, 161–176 (2012).
23. Z. Qian, P. Brunovskis, F. Rauscher 3rd, L. Lee, H. J. Kung, *J. Virol.* **69**, 4037–4044 (1995).
24. Materials and methods are available as supplementary materials.
25. S. R. Fiddaman *et al*., antonisdim, antonisdim/MDV: ancient MDV version 1.0.1, Zenodo (2023); doi:10.5281/zenodo.10022436.

ACKNOWLEDGMENTS

This research used the University of Oxford's Advanced Research Computing, Queen Mary's Apocrita, and the Leibniz-Rechenzentrum (LRZ) High Performance Computing facility. **Funding:** This work was supported by the European Research Council (grants ERC-2019-StG-853272-PALAEOFARM or ERC-2013-StG-337574-UNDEAD or both to S.R.F., L.A.F.F., G.L., and A.L.S.); Wellcome Trust (grant 210119/Z/18/Z to S.R.F. and L.A.F.F.); the Oxford Martin School Pandemic Genomics Programme (to S.R.F., L.d.P., and O.G.P.); AHRC (grant AH/L006979/1 to G.L., O.L., and N.S.); European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie (grant agreement 895107 to O.L.); BBSRC (grant BB/M011224/1 to S.D.); and Postdoctoral grant 12U7121N of the Research Foundation–Flanders (Fonds voor Wetenschappelijk Onderzoek) (to B.V.). **Author contributions:** Conceptualization: S.R.F., A.L.S., L.A.F.F., and G.L. Methodology: S.R.F., E.A.D., A.L.S., L.A.F.F., G.L., B.V., L.d.P., V.N., O.L., and O.G.P. Sample provision: O.L., N.M., G.F., R.S., H.B., L.D.-S., D.N.S., I.V.A., O.P., M.S., H.D., H.F., A.S.M., A.A.V., A.F., N.S., J.B., A.O.A., O.V.A., M.M., L.Z., and V.N. Investigation: S.R.F., E.A.D., O.L., L.d.P., B.V., S.C., A.F.H., K.T., P.G.F., S.D., H.L., G.C.B., O.G.P., V.N., G.L., A.L.S., and L.A.F.F. Visualization: S.R.F. Funding acquisition: L.A.F.F., A.L.S., and G.L. Project administration: S.R.F., L.A.F.F., A.L.S., and G.L. Supervision: L.A.F.F., A.L.S., and G.L. Writing – original draft: S.R.F., L.A.F.F., A.L.S., and G.L. Writing – review and editing: S.R.F., E.A.D., O.L., L.d.P., B.V., S.C., A.F.H., K.T., P.G.F., S.D., N.M., H.L., G.F., R.S., H.B., L.D.-S., D.N.S., I.V.A., O.P., M.S., H.D., H.F., A.S.M., A.A.V., A.F., N.S., G.C.B., J.B., A.O.A., O.V.A., M.M., O.G.P., V.N., G.L., A.L.S., and L.A.F.F. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All MDV sequence data generated have been deposited in GenBank under accession PRJEB64489. Code is available at GitHub (https://github.com/antonisdim/MDV) and archived at Zenodo (https://zenodo.org/records/10022436) (*25*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by Wellcome Trust (210119/Z/18/Z), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg2238
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S13
References (*26*–*74*)
MDAR Reproducibility Checklist

Submitted 22 December 2022; resubmitted 15 March 2023
Accepted 25 October 2023
10.1126/science.adg2238

CONSERVATION

Genomic analyses reveal poaching hotspots and illegal trade in pangolins from Africa to Asia

Jen C. Tinsman[1,2,3,4]*, Cristian Gruppi[1,3], Christen M. Bossu[5], Tracey-Leigh Prigge[6,4], Ryan J. Harrigan[1,3], Virginia Zaunbrecher[1,3], Klaus-Peter Koepfli[7,8], Matthew LeBreton[3,9,10], Kevin Njabo[1,3], Cheng Wenda[11,6], Shuang Xing[11,6], Katharine Abernethy[12,13], Gary Ades[14], Excellence Akeredolu[15], Imuzei B. Andrew[15], Taneisha A. Barrett[6], Iva Bernáthová[16], Barbora Černá Bolfíková[16], Joseph L. Diffo[17], Ghislain Difouo Fopa[18,4], Lionel Esong Ebong[19], Ichu Godwill[20,4], Aurélie Flore Koumba Pambo[21], Kim Labuschagne[22], Julius Nwobegahay Mbekem[23], Brice R. Momboua[21,24], Carla L. Mousset Moumbolou[21,24,25,4], Stephan Ntie[24,21], Elizabeth Rose-Jeffreys[14], Franklin T. Simo[18,4], Keerthana Sundar[1], Markéta Swiacká[26], Jean Michel Takuo[10,17], Valery N. K. Talla[27,28], Ubald Tamoufe[17], Caroline Dingle[6], Kristen Ruegg[5], Timothy C. Bonebrake[6,3,4]*, Thomas B. Smith[1,3,29]*

The white-bellied pangolin (*Phataginus tricuspis*) is the world's most trafficked mammal and is at risk of extinction. Reducing the illegal wildlife trade requires an understanding of its origins. Using a genomic approach for tracing confiscations and analyzing 111 samples collected from known geographic localities in Africa and 643 seized scales from Asia between 2012 and 2018, we found that poaching pressures shifted over time from West to Central Africa. Recently, Cameroon's southern border has emerged as a site of intense poaching. Using data from seizures representing nearly 1 million African pangolins, we identified Nigeria as one important hub for trafficking, where scales are amassed and transshipped to markets in Asia. This origin-to-destination approach offers new opportunities to disrupt the illegal wildlife trade and to guide anti-trafficking measures.

Overexploitation is one of the greatest threats to biodiversity worldwide (*1*). In particular, the illegal wildlife trade is accelerating the extinction of thousands of species globally (*2*, *3*). Determining where animals are poached is a major challenge in curbing this exploitation, especially for internationally trafficked species such as pangolins (*4*–*7*). Pangolin scales are used as an ingredient in traditional medicines, despite no evidence of their efficacy (*8*). The largest markets for these products are in China (*9*–*11*). As populations of the Asian species of pangolins have declined, smugglers have begun importing African pangolins to meet demand (*5*, *12*–*14*). Consequently, the white-bellied pangolin (*Phataginus tricuspis*), which occurs throughout Western and Central Africa, from Guinea to Zambia, is now the most trafficked mammal in the world (*11*, *15*–*19*).

Given the extensive geographic range of white-bellied pangolins and many other trafficked species, identifying poaching hotspots at a scale that is useful to law enforcement is crucial for conservation efforts. Without genetic data to reveal the true geographic origins of poached animals or products, seizures by law enforcement agencies offer limited information about their sources. For example, pangolin scales arriving in China from Nigeria may have originated there or been amassed and transshipped from other countries (*5*, *20*). Determining the precise origins of animals involved in the global wildlife trade is an urgent priority (*3*, *21*).

Here, we report an origin-to-destination approach for understanding where pangolins are harvested, amassed, shipped, and consumed. First, we mapped geographically and genetically distinct populations of white-bellied pangolin using 111 samples collected from wild pangolins at known localities across their range. We then used this spatially explicit genomic map to assign 643 confiscated scales to their previously unknown geographic origins (*22*). By sampling 32 confiscations seized between 2012 and 2018, representing over 100,000 trafficked pangolins, we mapped pangolin poaching hotspots [illegible] documented changes in poaching pressure [illegible] time. Next, we analyzed data on pangolin trafficking incidents to connect poaching hotspots with common trafficking routes to markets. Our approach enables the monitoring of changes in poaching in near real-time, allowing for targeted and more effective anti-poaching measures.

Building a spatially explicit genomic map for white-bellied pangolins

We obtained 551 georeferenced samples from wild white-bellied pangolins. These samples included blood dots, muscle, and scales donated by pangolin hunters and recent tissue specimens from natural history collections (see materials and methods and data S1). To ensure fine-scale resolution for the map, we only included samples whose locality data were collected with a GPS unit. We excluded samples from urban and suburban bush meat markets, which may have obtained their pangolins through regional supply chains.

We extracted DNA that yielded high-quality genome sequences from 111 samples, which were used to construct the white-bellied pangolin genomic map (Fig. 1). The analyzed genomes contained over 4 million single-nucleotide polymorphisms (SNPs). We used principal component analysis and ADMIXTURE (*23*) to visualize genetic variation and identify distinct genetic clusters (fig. S1). We identified five distinct population groups that were strongly associated with distinct geographic regions (Fig. 1).

Genotyping confiscated scales to reveal poaching hotspots

We used the genomic variation detected in wild populations to generate a panel of 96 diagnostic SNPs for assigning unknown samples to their geographic origins (data S2). We tested the accuracy and precision of the SNP genotyping assay using the 111 mapped samples by first assigning them to distinct genetic clusters using the program rubias (*24*) and then predicting the geographic location of samples using OriGen (*25*). Rubias assigned 110 samples to the correct genetic cluster (99% accuracy), and our assay correctly localized 87.4% of these samples using OriGen (defined as within 500 km of

[1]Center for Tropical Research, Institute of the Environment, University of California, Los Angeles, Los Angeles, CA, USA. [2]National Fish and Wildlife Forensic Laboratory, US Fish and Wildlife Service, Ashland, OR, USA. [3]Congo Basin Institute, University of California, Los Angeles, Los Angeles, CA, USA. [4]Pangolin Specialist Group, IUCN Species Survival Commission, London, UK. [5]Department of Biology, Colorado State University, Fort Collins, CO, USA. [6]School of Biological Sciences, The University of Hong Kong, Hong Kong, China. [7]Smithsonian-Mason School of Conservation, George Mason University, Front Royal, VA, USA. [8]Center for Species Survival, Smithsonian's National Zoo and Conservation Biology Institute, Washington, DC, USA. [9]Mosaic, Yaoundé, Cameroon. [10]International Institute for Tropical Agriculture, Yaoundé, Cameroon. [11]School of Ecology, Shenzhen Campus of Sun Yat-sen University, Shenzhen, China. [12]Institut de Recherche en Ecologie Tropicale, Centre National de la Recherche Scientifique et Technologique, Libreville, Gabon. [13]Biological and Environmental Sciences, University of Stirling, Stirling, UK. [14]Fauna Conservation Department, Kadoorie Farm and Botanic Garden, Hong Kong, China. [15]Department of Zoology, Faculty of Science, University of Lagos, Lagos, Nigeria. [16]Faculty of Tropical AgriSciences, Czech University of Life Sciences Prague, Prague, Czech Republic. [17]Metabiota Cameroon Ltd, Yaoundé, Cameroon. [18]Department of Biology and Animal Physiology, University of Yaoundé I, Yaoundé, Cameroon. [19]Department of Ecology and Nature Management, School of Earth Sciences and Environmental Engineering, National Research Tomsk Polytechnic University, Tomsk, Russia. [20]Department of Wildlife, Fisheries and Aquaculture, Mississippi State University, Starkville, MS, USA. [21]Agence Nationale des Parcs Nationaux, Libreville, Gabon. [22]South African National Biodiversity Institute, Pretoria, South Africa. [23]CRESAR, Yaoundé, Cameroon. [24]Département de Biologie, Faculté des Sciences, Université des Sciences et Techniques de Masuku, Franceville, Gabon. [25]Pangolin Conservation Network, Libreville, Gabon. [26]Department of Spatial Sciences, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic. [27]Département de Biologie des Organismes, Université Libre de Bruxelles, Brussels, Belgium. [28]Laboratory of Applied Biology and Ecology, Faculty of Science, University of Dschang, Dschang, Cameroon. [29]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, USA.
*Corresponding author. Email: tinsman@ucla.edu (J.C.T.); tbone@hku.hk (T.C.B.); tbsmith@ucla.edu (T.B.S.)

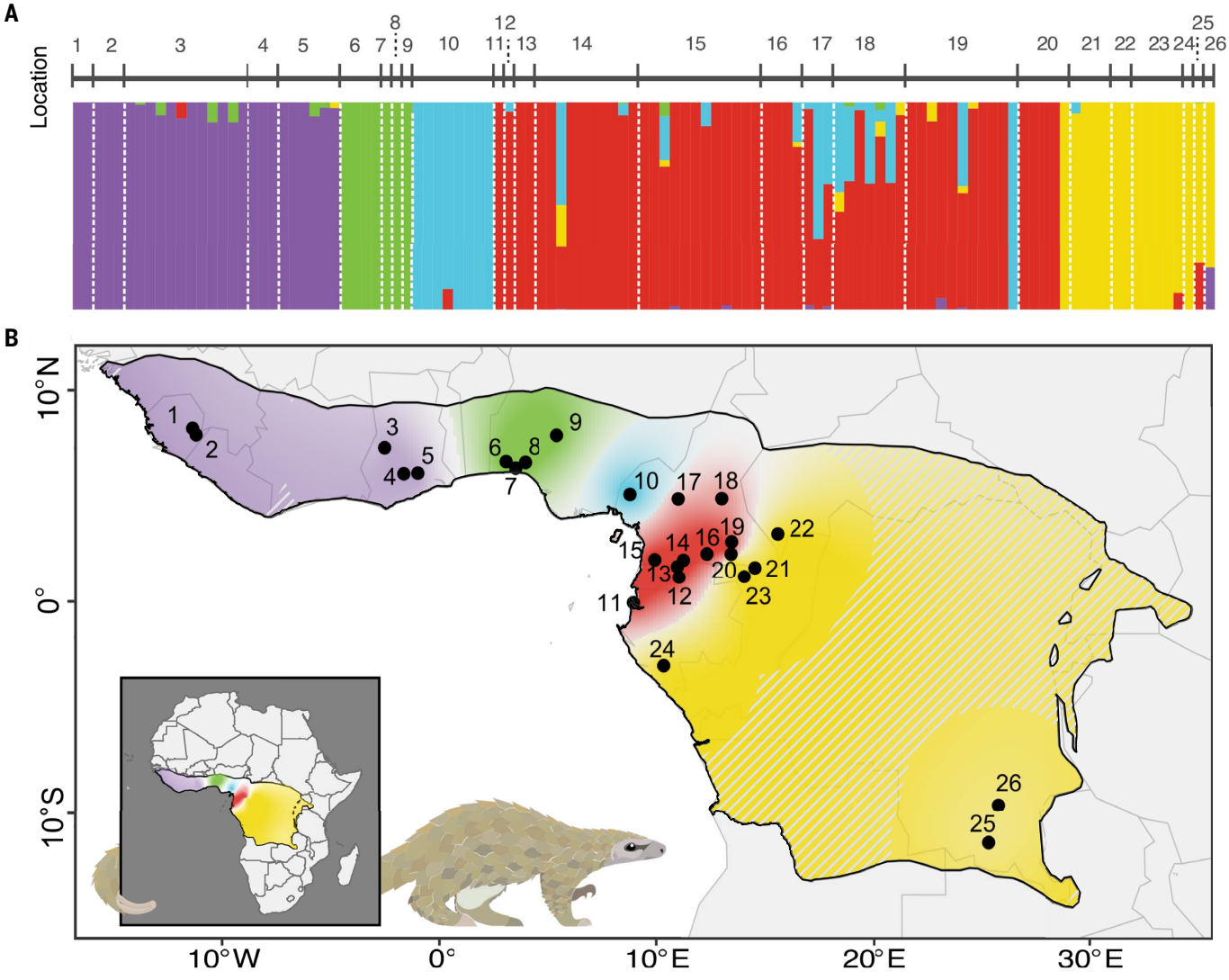


Fig. 1. Map of genomically identified white-bellied pangolin population clusters. (**A**) ADMIXTURE plot based on whole genomic data. The colored bars represent each sample's probability of assignment to genetically distinct population clusters. (**B**) Map of population clusters. Black dots indicate the 26 locations of the 111 samples included in the construction of the map. The intensity of the color represents the interpolation model's confidence that the indicated cluster occurs in that location. Areas more than 500 km away from a sampled locality are striped, showing more extensive interpolation, and will likely reveal subpopulation structure with additional sampling (fig. S1). [Pangolin illustration by S. McCabe]

their known origin; fig. S2). The 97 correct localizations had a median error of 63.5 km (mean, 126.0 km). Of the 14 incorrectly localized samples, 13 were still assigned to the correct genetic cluster. Median error for all scale location assignments was 72.2 km; mean error was 236.1 km.

Assignment errors were not randomly distributed across population clusters, either in magnitude or in frequency (fig. S2 and tables S1 to S3). We achieved the best resolution in the population clusters where sampling was densest. Incorrect localizations happened in clusters of populations spanning large geographic areas with fewer samples. For instance, the assay predicted that two individuals from southeastern Democratic Republic of the Congo (DRC) originated closer to the Central African Republic (fig. S2).

Additional sampling likely will resolve subpopulations within the five major population clusters (Fig. 1). This might also reduce the frequency and magnitude of misassignment errors, especially in the West African and Congo Basin clusters. Until the SNP assay is refined to include data from undersampled regions, OriGen-generated coordinates that fall within those ranges or plot far outside of densely sampled areas should be treated with caution (*25*).

We SNP genotyped 656 confiscated white-bellied pangolin scales. Authorities in Hong Kong SAR, China, seized these scales from 32 shipments, arriving from at least seven separate transit routes, between 2012 and 2018 (data S3). These seizures weighed a collective 38 tonnes and represented an estimated minimum of 105,447 dead pangolins. We successfully genotyped 647 of these scales and assigned 643 (98%) to one of five distinct pangolin genetic clusters with >80% posterior probability. We then estimated their geographic origins (fig. S3).

Genotyping confiscated white-bellied pangolin scales revealed two major pangolin poaching hotspots. The vast majority of the 643 genotyped scales originated along Cameroon's southern border with Equatorial Guinea and Gabon, and from western Cameroon, near the border with Nigeria (Fig. 2 and data S4).

Comparing genomics with data on pangolin seizures

To compare genetically determined pangolin origins to those available from extant data sources, and to identify the trafficking routes used to smuggle scales to markets, we developed

a database of pangolin seizures (data S5). We found data on 3097 seizures of pangolins from January 1981 through December 2022 (fig. S4), at least 999 of which included African species. Using published scale-mass-to-individual conversion metrics (*16*, *20*), we estimated that these seizures represented at least 986,894 poached African pangolins.

Network mapping of African pangolin seizures identified Nigeria as the highest-volume transit hub in Africa, where traffickers amass pangolin scales before shipping them overseas (Fig. 3 and fig. S5) (*20*, *26*). Nigeria's seizures are more comparable to those of non-African transit locations, such as Hong Kong SAR, China, and Turkey (table S4). However, data from public records of illegal activity are subject to major biases such as law enforcement effort and media interest in large seizures (*27*). We investigated whether these reports could provide a reliable picture of where pangolins are poached.

The seizures database rarely indicated non-Nigerian origins for pangolins that transited through Nigeria (95.1% of seized animals have no recorded source other than Nigeria). By contrast, our genetic assay results show that only 4.2% of pangolins shipped from Nigeria originated there. Most of the pangolins in our samples that transited through Nigeria originated in southern Cameroon, mainland Equatorial Guinea, and Gabon (Fig. 3). For example, a 2018 seizure consisted of 7.1 tonnes of scales shipped by cargo container from Nigeria to Hong Kong SAR, China (#613/2018, data S3). However, the pangolins in our sample of this seizure originated in Cameroon, Equatorial Guinea, Gabon, and the Republic of the Congo (fig. S6; data S4).

Our results reveal the importance of using genetic data to understand pangolin poaching. Reports by law enforcement officials and anti-trafficking nongovernmental organizations such as those in the seizures database can show major shipping routes (figs. S7 to S9) and guide the interception of shipments already in progress, but they cannot reliably identify where pangolins are harvested (Fig. 3).

Conversely, genetic assignments of scales do not reveal the intermediate stops that they take to market. Taken together, genetic analyses and seizure data show a major origin-to-destination trafficking route for white-bellied pangolins. Results suggest that the samples that we analyzed were harvested in southern Cameroon, Equatorial Guinea, and Gabon; amassed in Nigeria; transported to intermediate destinations in southeast Asia, often by sea; and ultimately used in southeastern provinces of mainland China, particularly Guangdong and Guangxi (Fig. 4).

Tracking poaching over time

Our results provide a geographically explicit understanding of where global trafficking networks threaten white-bellied pangolins most. However, our study and past work suggest that these routes will likely change over time, and continuous monitoring is necessary to detect changes in trafficking patterns (*18*).

Indeed, testing seizures confiscated over just 7 years (2012–2018) enabled us to detect changes in the origins of trafficked pangolins. Early on, poaching activity was confined to West Africa before shifting to Central Africa more recently (Fig. 5). These changes in trafficking patterns could represent (i) a response to increased enforcement; (ii) declining pangolin populations in West Africa; or (iii) taking advantage of new, convenient trade routes, or a combination thereof (*14*, *16*). With over half a million African pangolins seized from the illegal wildlife trade in that time frame (fig. S4), unsustainable exploitation seems all but certain. We believe this number probably represents a gross underestimate of the trade in African species, because most confiscations in Asia do not document species or origin information. Further, many shipments are never detected at all (*16*).

Unsustainable harvest of the West African population of white-bellied pangolin has shifted pressure onto two geographically restricted populations (Figs. 1 and 5). Their limited range and high levels of exploitation make them some of the most threatened populations of white-bellied pangolins. Moreover, the threats that these populations face will likely increase with the construction of new ports, roads, and rail lines in the region (*28*–*30*). For example, the rapidly growing Kribi Deepwater Port facility and associated road network in southwestern Cameroon could easily provide transport for pangolins poached in the southern hotspot that we identified (Fig. 2). As a result, populations in southern Cameroon, northern Equatorial Guinea, and Gabon are at particularly high risk of overexploitation (*30*–*32*).

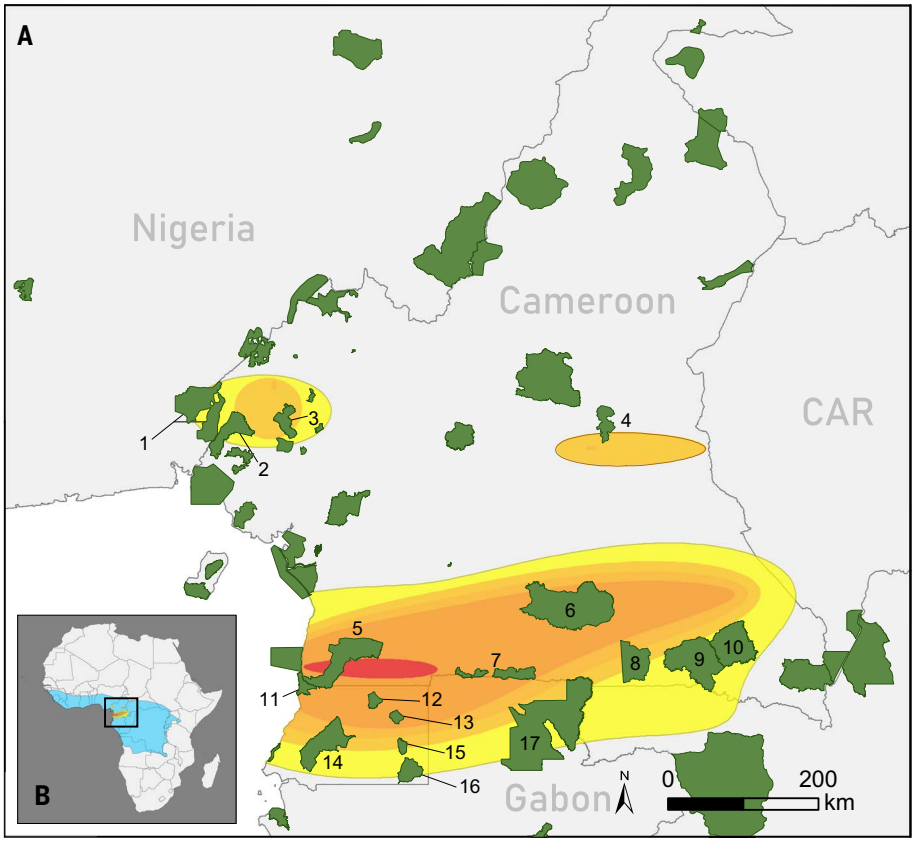


Fig. 2. Pangolin poaching hotspots. (**A**) Hotspots were derived from scales confiscated in Hong Kong SAR, China, between 2012 and 2018. Warm colors indicate areas where the bulk of scales in our sample originated, and warmer colors indicate an even greater density of scales originating there. Protected areas (in green) associated with poaching hotspots are (1) Cross River National Park, Nigeria; (2) Korup National Park, Cameroon; (3) Bayang Mbo Wildlife Sanctuary, (4) Deng Deng National Park, (5) Campo Ma'an National Park, (6) Dja Faunal Reserve, (7) Mengame Gorilla Sanctuary, (8) Ngoyla Faunal Reserve, (9) Nki National Park, (10) Boumba Bek National Park, and (11) Rio Campo Nature Reserve, Equatorial Guinea; (12) Monte Temelón Nature Reserve, (13) Piedra Bere Natural Monument, (14) Monte Alén National Park, (15) Piedra Bere Natural Monument, (16) Altos de Nsork National Park, and (17) Minkébé National Park, Gabon. Protected Area was defined as any conservation landscape falling into IUCN Protected Area Categories I–IV (UNEP-WCMC and IUCN 2022). (**B**) Location of hotspots in relation to the range of the white-bellied pangolin (in blue) (*33*).

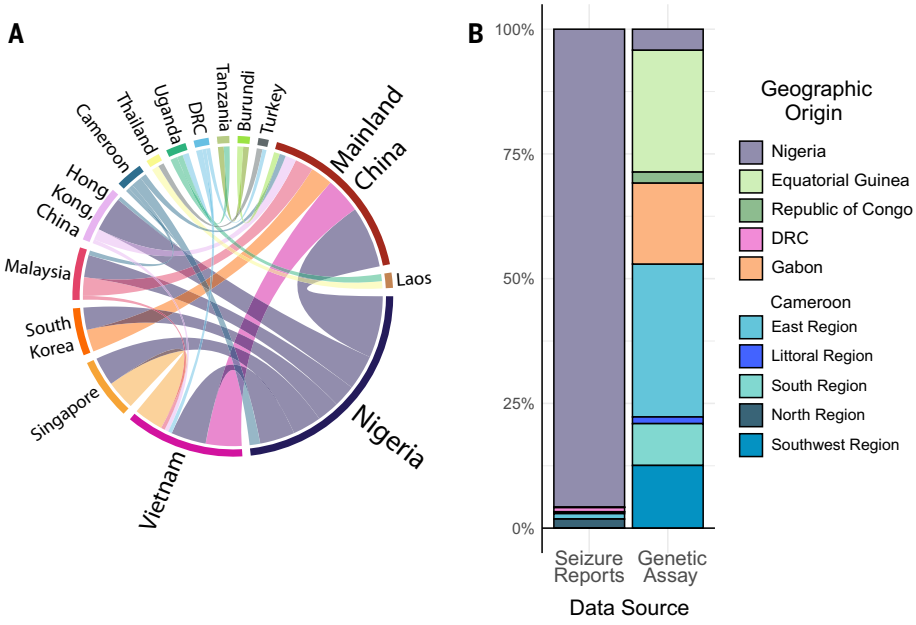


Fig. 3. Network model of the flow of African pangolins and a comparison of genetic and report-based data. (**A**) Line widths indicate the relative quantity of pangolins shipped along each route, and line colors indicate the origin country. Only routes with a total greater than 10,000 pangolins and/or individual-equivalents by weight shipped are shown. (**B**) The origins of smuggled pangolins that transited through Nigeria, as indicated by traditional data from seizures reports and as determined using genomics. Percentages are also presented in tables S5 and S6.

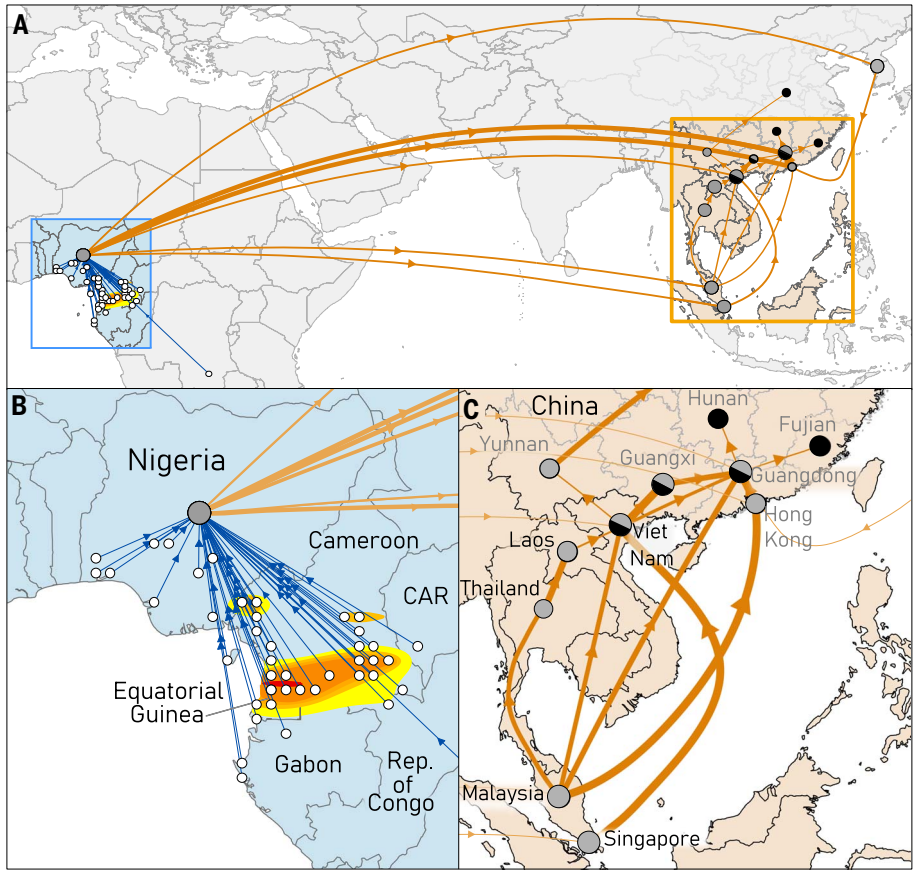


Fig. 4. An origin-to-destination map of pangolin trafficking. A combination of genomics (blue lines) and publicly reported data on pangolin seizures (orange) reveal major trafficking routes. This map focuses on pangolins that transited through Nigeria. White dots represent estimated pangolin origins, transit locations are gray, and market or consumption locations are black. Line widths reflect the quantity of pangolins smuggled along a route. These lines represent possible routes between known stops, not the actual paths taken by trafficked scales. (**A**) Transcontinental routes for trafficking African pangolins to Asia. CAR, Central African Republic. (**B**) A zoomed-in look at source localities for African pangolins transited through Nigeria. We picked a central point in Nigeria for visualization—most of these scales left the country via seaport. (**C**) Routes taken by African pangolins once they arrive in Southeast Asia.

When wildlife traffickers encounter increased enforcement, they typically move their operations (*18*). The genetic assay presented here will allow for near real-time monitoring of shifts in poaching hotspots. Our assay can localize most samples within 100 km of their geographic origin (median error: 72.2 km across all samples). This is very precise, considering that the white-bellied pangolin's distribution covers roughly 6 million km^2 (*33*, *34*).

Large confiscations of white-bellied pangolin scales often represent multiple individuals poached in different locations. If sampled forensically, these seizures could give us a snapshot of poaching activity across West and Central Africa, complementing and extending data available from elephant ivory seizures (*6*, *7*, *34*). Given the discovery of novel severe acute respiratory syndrome (SARS)–related coronaviruses in Asian pangolin seizures (*35*), our assay could also provide spatial insights

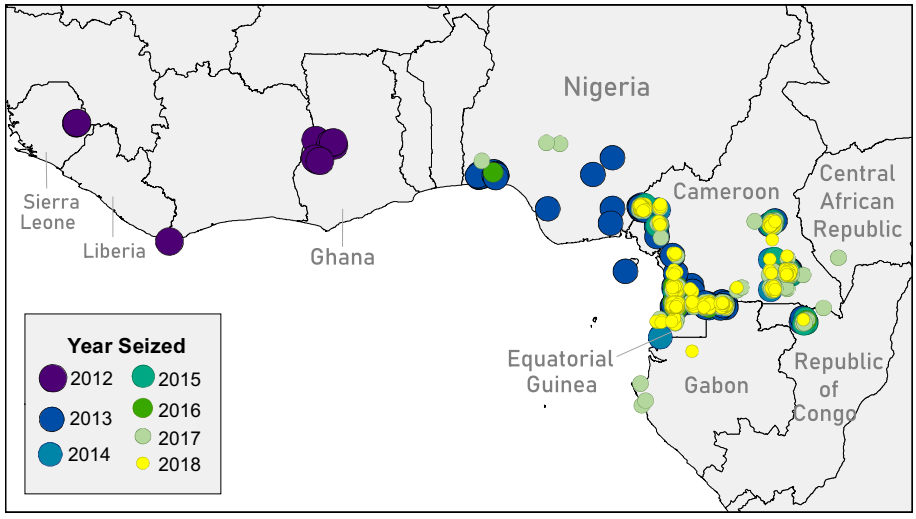


Fig. 5. Changes in pangolin poaching over time. A map of the predicted origins of 643 confiscated white-bellied pangolin scales. Authorities in Hong Kong SAR, China, confiscated these scales in 32 shipments originating from Nigeria, Cameroon, Egypt, Cote d'Ivoire, and Kenya (data S3). One scale, which originated in southeastern DRC and was seized in 2017, is not pictured. The precise location of the scales that were assigned to the West African population cluster (purple), should be treated with caution, as the SNP assay currently experiences relatively high rates of localization error in that cluster. Even so, given the diversity of genotypes observed in the scales from 2012, it seems likely that there were multiple geographic sources within the West African cluster.

into zoonotic disease risk. Compared with traditional law enforcement investigations, the genetic assay reduces the time lag between intercepting wildlife products, tracing an international supply chain to its origins, and reactive enforcement. This approach can dynamically guide preventive efforts by revealing poaching hotspots, representing an important step forward in conserving this highly trafficked species.

REFERENCES AND NOTES

1. S. L. Maxwell, R. A. Fuller, T. M. Brooks, J. E. M. Watson, *Nature* **536**, 143–145 (2016).
2. UNODC, "World Wildlife Crime Report: Trafficking in protected species" (2020).
3. A. C. Hughes, *Curr. Biol.* **31**, R1218–R1224 (2021).
4. D. W. S. Challender, S. R. Harrop, D. C. MacMillan, *Glob. Ecol. Conserv.* **3**, 129–148 (2015).
5. W. Cheng, S. Xing, T. C. Bonebrake, *Conserv. Lett.* **10**, 757–764 (2017).
6. S. K. Wasser *et al.*, *Conserv. Biol.* **22**, 1065–1071 (2008).
7. S. K. Wasser *et al.*, *Nat. Hum. Behav.* **6**, 371–382 (2022).
8. R. L. Jacobs, P. J. McClure, B. W. Baker, E. O. Espinoza, *Conserv. Sci. Pract.* **1**, e82 (2019).
9. D. W. S. Challender, *Traffic Bull.* **23**, 92–93 (2011).
10. F. Hornor, D. Thorne, A. Shaver, "Tipping the scales: Exposing the growing trade of African pangolins into China's traditional medicine industry" [C4ADS and USAID Reducing Opportunities for Unlawful Transport of Endangered Species (ROUTES), 2020], pp. 1–60.
11. Z.-M. Zhou, Y. Zhou, C. Newman, D. W. Macdonald, *Front. Ecol. Environ.* **12**, 97–98 (2014).
12. D. W. S. Challender, L. Hywood, *Traffic Bull.* **24**, 53–55 (2012).
13. L. Gomez, B. T. C. Leupen, T. K. Hwa, *Traffic Bull.* **28**, 3–5 (2016).
14. M. M. Mambeya *et al.*, *Afr. J. Ecol.* **56**, 601–609 (2018).
15. D. W. S. Challender, C. Waterman, J. E. M. Baillie, "Scaling up pangolin conservation: IUCN SSC Pangolin Specialist Group conservation action plan" (Zoological Society of London, 2014); https://portals.iucn.org/library/sites/library/files/documents/2014-062.pdf.
16. D. W. S. Challender, S. Heinrich, C. R. Shepherd, L. K. D. Katsis, in *Pangolins*, D. W. S. Challender, H. C. Nash, C. Waterman, Eds. (Academic Press, 2020), pp. 259–276.
17. K. M. Ewart *et al.*, *Forensic Sci. Int. Anim. Environ.* **1**, 100014 (2021).
18. S. Heinrich *et al.*, *The Global Trafficking of Pangolins: A Comprehensive Summary of Seizures and Trafficking Routes from 2010–2015* (TRAFFIC, 2017).
19. D. J. Ingram *et al.*, *Conserv. Lett.* **11**, e12389 (2018).
20. C. A. Emogor *et al.*, *Biol. Conserv.* **264**, 109365 (2021).
21. D. W. Pietersen, D. W. S. Challender, in *Biodiversity of World: Conservation from Genes to Landscapes*, D. W. S. Challender, H. C. Nash, C. Waterman, Eds. (Academic Press, 2020), pp. 537–543.
22. K. C. Ruegg *et al.*, *Mol. Ecol.* **23**, 5726–5739 (2014).
23. D. H. Alexander, J. Novembre, K. Lange, *Genome Res.* **19**, 1655–1664 (2009).
24. E. C. Anderson, rubias: Bayesian inference from the conditional genetic stock identification mode (2022); http://ftp.edu.ee/pub/cran/web/packages/rubias/rubias.pdf.
25. J. M. Rañola, J. Novembre, K. Lange, *Bioinformatics* **30**, 2915–2922 (2014).
26. M. Utermohlen, P. Baine, "In Plane Sight: Wildlife Trafficking in the Air Transport Sector - Wildlife Trade Report from TRAFFIC" [C4ADS and USAID Reducing Opportunities for Unlawful Transport of Endangered Species (ROUTES), 2018]; https://www.traffic.org/site/assets/files/10858/in_plane_sight.pdf.
27. F. M. Underwood, R. W. Burn, T. Milliken, *PLOS ONE* **8**, e76539 (2013).
28. A. C. Hughes *et al.*, *Trends Ecol. Evol.* **35**, 583–593 (2020).
29. H. Yang *et al.*, *Nat. Ecol. Evol.* **5**, 1520–1529 (2021).
30. W. F. Laurance, S. Sloan, L. Weng, J. A. Sayer, *Curr. Biol.* **25**, 3202–3208 (2015).
31. W. E. Laurance *et al.*, *Conserv. Biol.* **20**, 1251–1261 (2006).
32. D. Wilkie, E. Shaw, F. Rotberg, G. Morelli, P. Auzel, *Conserv. Biol.* **14**, 1614–1622 (2000).
33. D. W. Pietersen *et al.*, The IUCN Red List of Threatened Species. *White-Bellied Pangolin Phataginus Tricuspis* (2019); https://doi.org/10.2305/IUCN.UK.2019-3.RLTS.T12767A123586469.en.
34. S. K. Wasser *et al.*, *Science* **349**, 84–87 (2015).
35. T. T.-Y. Lam *et al.*, *Nature* **583**, 282–285 (2020).
36. J. C. Tinsman *et al.*, Genomic analyses reveal poaching hotspots and illegal trade in pangolins from Africa to Asia (2023); https://doi.org/10.5061/dryad.zkh1893g7.

ACKNOWLEDGMENTS

We thank the Agriculture, Fisheries and Conservation Department of the Hong Kong SAR Government and Kadoorie Farm and Botanic Garden for donating confiscated pangolin scales to this project. We thank the American Museum of Natural History, Metabiota, Sangha Pangolin Project, and the South African National Biodiversity Institute for donating samples. We thank the Environmental Investigation Agency and TRAFFIC's Wildlife Trade Portal for sharing their seizures data. For permission to collect samples, we thank the Ministry of Scientific Research and Innovation and the Ministry of Forests and Wildlife of Cameroon; Centre National de la Recherche Scientifique et Technologique and Agence Nationale des Parcs Nationaux of Gabon; and Ministère de la Recherche Scientifique et de l'Innovation Technologique and Agence Congolaise de la Faune et des Aires Protégées of the Republic of the Congo. We thank African Parks Network and the Congo Basin Institute for their logistical support. We thank S. Wasser for suggestions on our approach. We are grateful to P. Ricca and P. Tonnis for help organizing sample shipments. J.C.T. thanks J. Eld, T. Alp, H. Omer, and M. Ikizler, and E. Tinsman for childcare, which made this work possible. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the US Fish and Wildlife Service. This article was funded in part by a grant from the United States Department of State. The opinions, findings, and conclusions stated herein are those of the authors and do not necessarily reflect those of the United States Department of State. **Funding:** Bureau of International Narcotics and Law Enforcement Affairs, US Department of State S-INLEC-17-GR-1006 (T.B.S.); National Geographic Society NGS-418C-18 (T.B.S.), 7762-04 (M.L.); National Science Foundation Postdoctoral Research Fellowship in Biology DBI 2208955 (J.T.); Research Impact Fund, Research Grants Council Hong Kong R7021-20 (T.C.B., C.D.); National Institutes of Health Director's Pioneer Award DP1-OD000370 and EU-ACP ECOFAC VI Convention FED/2018/403-718 (A.F.K.P., B.R.M.); Czech Ministry of Interior VK01010103 (B.C.B.); IGA Faculty of Environmental Sciences CZU Prague 2021B0026 (MS); IDEA WILD EBONCAME0221 (LEE). Funding for sampling in Cameroon: google.org, Skoll Foundation, United States Agency for International Development (USAID) Emerging Pandemic Threats PREDICT program (GHN-AOO-09-00010-00) **Author contributions:** Conceptualization: R.J.H., V.Z., M.L., K.N., K.R., T.C.B., T.B.S.; Data Curation: J.C.T., C.G., C.M.B., T.-L.P., R.J.H., C.W., S.X.; Formal Analysis: J.C.T., C.G., C.M.B., T.-L.P., R.J.H., C.W., S.X., T.B.S.; Funding Acquisition: J.C.T., R.J.H., V.Z., K.-P.K., M.L., K.N., K.A., L.E.E., A.F.K.P., B.R.M., C.D., T.C.B., T.B.S.; Investigation: all authors; Methodology: J.C.T., C.G., C.M.B., T.-L.P., R.J.H., M.L., C.W., S.X., T.C.B., T.B.S.; Project Administration: J.C.T., V.Z., M.L., K.N., K.R., T.C.B., T.B.S.; Resources: K.-P.K., M.L., K.A., G.A., I.B.A., T.A.B., J.L.D., L.E.E., G.D.F., I.G., A.F.K.P., K.L., B.R.M., C.L.M.M., S.N., J.N.M., E.R.J., F.T.S., K.S., M.S., J.M.T., V.N.K.T., U.T., T.C.B., T.B.S.; Supervision: C.G., V.Z., M.L., K.A., E.A., I.B., B.Č.B., A.F.K.P., C.D., K.R., T.C.B., T.B.S.; Visualization: J.C.T., C.G., C.M.B., T.-L.P., R.J.H., C.W., S.X., T.B.S.; Writing – original draft: J.C.T., C.G., C.M.B., T.-L.P., R.J.H., T.B.S.; Writing – review and editing: all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text, the supplementary materials, the NCBI Sequence Read Archive (BioProject PRJNA1014914), or (*36*), except wildlife trafficking data owned by the Wildlife Trade Portal and the Environmental Investigation Agency. Those organizations make their data available at https://www.wildlifetradeportal.org/dashboard and https://eia-international.org/global-environmental-crime-tracker, respectively. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi5066
Materials and Methods
Figs. S1 to S14
Tables S1 to S8
References (*37*–*65*)
Data S1 to S5

Submitted 30 April 2023; accepted 10 November 2023
10.1126/science.adi5066

STELLAR ASTROPHYSICS

An observed population of intermediate-mass helium stars that have been stripped in binaries

M. R. Drout[1,2]*†, Y. Götberg[2]*†, B. A. Ludwig[1], J. H. Groh[3], S. E. de Mink[4,5], A. J. G. O'Grady[1,6], N. Smith[7]

The hydrogen-rich outer layers of massive stars can be removed by interactions with a binary companion. Theoretical models predict that this stripping produces a population of hot helium stars of ~2 to 8 solar masses ($M_\odot$), however, only one such system has been identified thus far. We used ultraviolet photometry to identify potential stripped helium stars then investigated 25 of them using optical spectroscopy. We identified stars with high temperatures (~60,000 to 100,000 kelvin), high surface gravities, and hydrogen-depleted surfaces; 16 stars also showed binary motion. These properties match expectations for stars with initial masses of 8 to 25 $M_\odot$ that were stripped by binary interaction. Their masses fall in the gap between subdwarf helium stars and Wolf-Rayet stars. We propose that these stars could be progenitors of stripped-envelope supernovae.

Approximately 70% of massive stars [initial masses of >8 solar masses ($M_\odot$)] interact with a binary companion during their lifetimes (*1*, *2*). Those binary interactions are expected to strip the hydrogen-rich envelopes from many massive stars, leaving an exposed hot and compact helium core. The resulting stripped stars have sufficiently long lifetimes to be observed and are expected to be numerous (*3*).

Binary-stripped massive stars are expected to influence multiple astrophysical processes: They are thought to be the progenitors of most hydrogen-poor core-collapse supernovae (*4*–*6*). The neutron stars that have been observed in gravitational wave events are thought to have undergone two phases of envelope stripping (*7*). And the high surface temperatures of stripped stars make them potential sources of ionizing photons (*8*, *9*).

Despite their predicted ubiquity, few binary-stripped helium stars with masses between ~2 and 8 $M_\odot$—which are expected to be produced by stars with initial masses between ~8 and 25 $M_\odot$—have been found. Many other types of hydrogen-deficient stars have been observed (*10*). These are classified as high-mass Wolf-Rayet (WR) stars (*11*), low-mass subdwarfs (*12*), extreme helium stars (*13*), and central stars of planetary nebulae (*14*), all of which have been found in binary systems (*15*–*17*). However, none of those classes occupy the mass range that has been predicted to produce most stripped-envelope supernovae or neutron star mergers (*7*). Only one hot helium star with an appropriate mass has been reported: the "quasi-WR" star in the system HD 45166 (*18*, *19*).

If such systems are truly rare, models of binary evolution would need to be revised. Alternatively, there could be an observational bias: The optical flux from intermediate-mass stripped stars might be hidden by a bright main sequence (MS) companion star. Although helium star mass-loss rates are uncertain (*20*), they are predicted to exhibit weaker wind features than luminous WR stars, so they could potentially have eluded detection in previous surveys targeting those features (*21*).

Ultraviolet photometry

Some stripped helium star binaries might be detectable by excess ultraviolet (UV) emission in their spectral energy distributions (*22*). To assess this possibility, we calculated synthetic spectra for a large set of hypothetical binaries containing a stripped star and an MS star (*20*). We find that many of the hypothetical systems remain obscured by the brightness of the MS star, but hot intermediate-mass helium stars paired with MS companions of $\lesssim$10 $M_\odot$ occupy a specific region of UV-optical color-magnitude diagrams (CMDs): blueward of the MS at intermediate luminosities of –1 mag > M_{UVM2} > –4 mag (where M_{UVM2} is the absolute magnitude in the UVM2 ultraviolet filter; figs. S4 and S5).

We searched for massive stars with UV magnitudes that fall within the CMD region predicted by our synthetic spectra. We targeted stars in the Large Magellanic Cloud (LMC) and Small Magellanic Cloud (SMC) galaxies, because they contain a large number of massive stars at known distances, with low obscuration by dust. We measured UV photometry using archival images from the Swift Ultraviolet Survey of the Magellanic Clouds (*23*). These images cover ~3 square degrees of the SMC and ~9 square degrees of the LMC in three UV filters at a resolution of 2.5 arcsec. To reduce the effects of crowding at that resolution, we used the forward modeling code The Tractor (*24*) to perform forced point-spread function photometry. We adopted the known locations of stars in the optical Magellanic Cloud Photometric Survey (*25*, *26*), which has better spatial resolution.

This process determined UV magnitudes for >500,000 sources in the directions of the LMC and SMC (*20*). Figure 1 shows a UV-optical CMD of all the sources. We adopted distances of 50 and 61 kpc and visual dust extinctions A_V of 0.38 and 0.22 mag for the LMC and SMC, respectively (*20*). LMC and SMC extinction curves (*27*) were used to determine the corresponding dust obscuration in the UV. The CMD contains a dense band (which we ascribe to the MS) and multiple sources blueward of the MS, which we consider to be candidate stripped helium star binaries.

Optical spectroscopy

We selected 25 candidate systems for follow-up spectroscopy by choosing targets that have luminosities and colors consistent with our predictions for binaries containing intermediate-mass helium stars (*20*) (indicated in Fig. 1). The stars are of similar brightness to MS stars with initial masses of ~6 to 15 $M_\odot$ but—for the adopted extinction—are located blueward of the zero-age MS (ZAMS) in nine distinct UV-optical CMDs (*20*). They have UV-optical colors similar to those of WR stars but are intrinsically fainter. For some systems, the observed colors and magnitudes approach predictions for isolated helium stars with masses between ~2 and 8 $M_\odot$ (Fig. 1).

We obtained between 1 and 30 optical spectra for each system using the Magellan Echellette spectrograph (*28*) on the 6.5-m Magellan Baade telescope at Las Campanas Observatory, Chile. All 16 systems with more than one epoch show radial velocity variations, consistent with being binary systems (table S9).

We used kinematics to reject any likely foreground objects. All 25 systems have average radial velocities consistent with expectations for stars in the LMC and SMC (*20*). We combined these with proper motions (*29*), finding that 23 systems have three-dimensional motion that is consistent with known O-type and B-type massive stars in the LMC and SMC (*20*) (O-type stars typically have initial masses of $\gtrsim$15 $M_\odot$, and B-type stars typically have initial masses of ~2 to 15 $M_\odot$). The remaining two objects (stars 5 and 6 in table S9) show slight offsets in proper motion but have data quality issues in the proper motion catalog. We therefore retained them in our sample.

Figure 2A shows examples of the spectra; the full sample is provided in figs. S16 to S21. We classify the stars into three broad groups:

[1]David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, Toronto M5S 3H4, Canada. [2]The Observatories of the Carnegie Institution for Science, Pasadena, CA 91101, USA. [3]Independent researcher, 2314 Leiden, Netherlands. [4]Max-Planck-Institut für Astrophysik, 85741 Garching, Germany. [5]Anton Pannekoek Institute for Astronomy, University of Amsterdam, 1090 GE Amsterdam, Netherlands. [6]Dunlap Institute for Astronomy and Astrophysics, University of Toronto, Toronto M5S 3H4, Canada. [7]Steward Observatory, University of Arizona, Tucson, AZ 85721, USA.
*Corresponding author. Email: maria.drout@utoronto.ca (M.R.D.); ygoetberg@carnegiescience.edu (Y.G.)
†These authors contributed equally to this work.

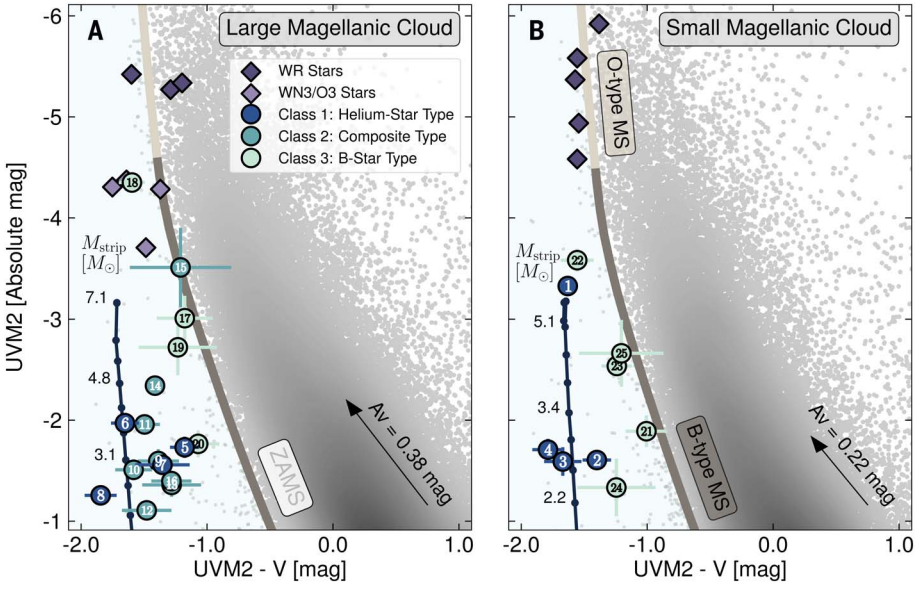


Fig. 1. Candidate stripped helium star binaries in UV-optical color-magnitude diagrams. Gray dots show absolute magnitude photometry in the UVM2 ultraviolet band as a function of the UVM2-V (where V is the visual band) color for stars in (**A**) the LMC and (**B**) the SMC. Numbered circles indicate the 25 stars we investigated further with optical spectroscopy (table S7), color coded according to their observed spectral morphologies (see legend). Error bars are 1σ. These systems have similar UV-optical colors, but lower brightnesses than either WR stars (dark-purple diamonds) (*48*, *49*) or the weaker-wind WN3/O3 stars (light-purple diamonds) (*36*). The connected black dots indicate models of isolated helium-core burning stripped stars, which are labeled with the current mass of the stripped star (M_{strip}) (*22*). The thick curved line indicates the expected position of the ZAMS for O-type (light gray) and B-type (dark gray) stars. All observed data have been corrected for dust extinction (indicated by the arrows), and all magnitudes are in the AB system.

1) Class 1 (eight stars): These spectra are dominated by absorption lines of He II. In some cases, lines of N IV and/or N V are visible in emission or absorption. These three ions are characteristic of very hot stars (*30*).

2) Class 2 (eight stars): These spectra show He II in absorption but also have strong absorption lines of hydrogen in the blue part of the spectrum (the short-wavelength Balmer lines). Six of these spectra also show He I; two do not, which indicates that the spectrum is a blend of two stars with different temperatures (*20*).

3) Class 3 (nine stars): No He II lines are visible in the optical spectrum. These spectra are dominated by strong Balmer and He I absorption lines, closely resembling those of B-type MS stars.

Interpretation of the spectra

The CMD locations and spectral morphologies of these 25 stars are consistent with our theoretical predictions for binary systems containing hot intermediate-mass helium stars. Our synthetic spectra of such objects (calculated above) show spectral features similar to those of WR stars, but with substantially weaker emission or absorption lines owing to their lower luminosities and mass-loss rates (*20*). In the set of helium star plus MS star composite spectral models, we identify the same three broad spectral classes as in our observations. Reproducing the absorption line spectra of the observed sample requires mass-loss rates for intermediate-mass helium stars that are at least an order of magnitude lower than extrapolations of WR mass loss (*20*), which is consistent with theoretical predictions (*31*).

We interpret the progression from class 3 to class 1 as an increasing contribution from the helium star to the optical flux of the system (*20*). Figure 2B compares the equivalent width (EW) of the He II λ5411 line as a function of Hη + He II λ3835 [where λ notation indicates the wavelengths in angstroms (Å), Hη is the seventh Balmer line, and + indicates blended lines] for both our observed sample and the composite models. The He II λ5411 line arises from the helium star; it is not expected at the cooler temperatures of B-type MS stars. We use the Hη + He II λ3835 blend to probe the presence of an MS companion, because strong short-wavelength Balmer lines are not expected in hot hydrogen-depleted stars [we have to use a blended line because our synthetic spectra show no isolated hydrogen lines in the optical range for these stars (*20*)]. In this parameter space, all 25 observed stars overlap with the predictions from the composite models, regardless of which helium star mass-loss rates are adopted (*20*).

For class 3 stars, we find 3σ upper limits on the EW of He II λ5411 of $\lesssim$0.2 Å (table S8). This corresponds to models where the helium star contributes <20% of the optical flux, and hence the spectra appear like those of B-type MS stars. Although such stars can show a UV excess, our models predict that they should be close to the ZAMS (as observed for the class 3 stars in Fig. 1). In contrast, the class 1 stars all have Hη + He II λ3835 EWs of <1.2 Å, consistent with models for systems where a helium star contributes >80% of the optical flux, so the spectra appear like those of isolated helium stars. The models only produce such spectra when the MS companion has a mass of $\lesssim 3.5\,M_{\odot}$. We infer that the class 1 objects either (i) come from binary systems where the two stars had very different initial masses; or (ii) have companions that are compact objects (neutron stars or black holes), not MS stars. Those systems would also have sufficiently blue colors to match the observations; in the CMD, they are close to models of isolated helium stars. Although dust extinction toward individual objects is uncertain, the location of these models in the CMD is similar to that of the class 1 objects (Fig. 1). The intermediate class 2 stars fall close to models where the helium star contributes 20 to 80% of the optical flux, so they have composite optical spectra with contributions from both stars.

Estimates of stellar properties

To assess the nature of the hot stars in these systems, we considered EW diagnostics that distinguish the optical spectra of stripped stars from MS stars and estimate their surface properties. We used the 1D nonlocal thermodynamic equilibrium radiative transfer code CMFGEN (*32*) to compute a set of spectral models with a range of effective temperatures (30 kK ≤ T_{eff} ≤ 100 kK, where kK is kilokelvin), surface gravities [4.0 < log(g/cm s^{-2}) < 6.0], and surface hydrogen mass fractions ($X_{H,surf}$ = 0.01, 0.1, 0.3, and 0.5, which are all depleted below MS values) (*20*). We choose a baseline mass-loss rate of $10^{-9} M_{\odot}$ year^{-1} because it produces absorption line spectra. We then tested the impact of this choice on our results by varying the assumed mass-loss rate by two orders of magnitude. These models cover a broad parameter space without making assumptions about the detailed evolutionary state of each system. Figure 3 shows these models compared with both the class 1 stars and O- and B-type MS model spectra (*33*, *34*) in three parameter spaces. We focused on the class 1 stars because we expect them to have minimal contamination from any companion.

To constrain the effective temperatures of these stars, Fig. 3A shows the EW of He II λ5411 as a function of the EW of He I λ5876, which provides a temperature diagnostic due to variations in the the helium ionization balance

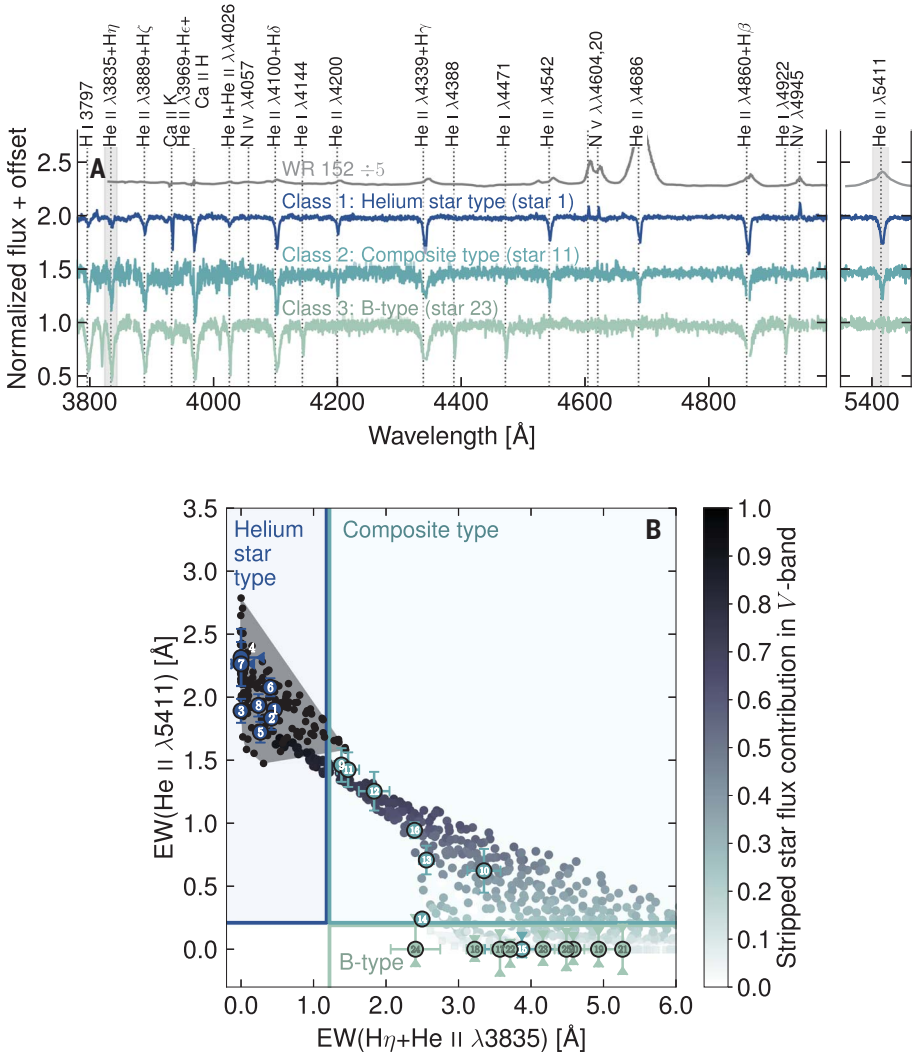


Fig. 2. Optical spectra with three spectral morphologies. (**A**) Three example observed spectra (colored lines) classified as class 1, 2, or 3 (see text), offset for display. Spectra of all the other stars in our sample are shown in figs. S16 to S21. The gray line shows the optical spectrum of an example WR star (WR 152 divided by a factor of 5) for comparison (*50*); it has similar line transitions as the class 1 stars, but in emission. Vertical dotted lines indicate locations of spectral lines, which are identified by the labels above. Gray shaded bands indicate the lines used in (B). (**B**) EWs of He II λ5411 and Hη + He II λ3835 for all 25 stars in our spectroscopic sample (large numbered circles). T-shaped error bars indicate 1σ uncertainties on detected lines; triangle-terminated error bars indicate 3σ upper or lower limits (for absorption or emission) on undetected lines. For comparison, we show synthetic models of single stripped stars (black dots enclosed in the gray shaded region), single B-type MS stars (light-green squares), and composites of the two (colored dots). Model equivalent widths were calculated assuming a signal-to-noise ratio of 35, consistent with the median signal-to-noise of the observed stars (*20*). Models are colored to indicate the fraction of *V*-band flux contributed to the binary by the stripped star (color bar). Shaded and labeled boxes define the three classes of spectral morphology we identify; observed data points use the same colors as the shading. Star 15 does not exhibit He II λ5411 but does show He II λ4686, so we classify it as class 2 (*20*). The observed sample forms a sequence that overlaps with the theoretical predictions for stripped helium star binaries.

with temperature. For all but two stars, we find $T_{\rm eff}$ > 70 kK owing to the lack of detected He I; those are temperatures typical of WR stars, higher than the hottest O-type stars (*35*). For some objects, the detection of N IV and/or N V can provide an alternative temperature estimate, which ranges between ~70 to 80 kK and ≳90 kK (fig. S11) (*20*).

Figure 3B shows an equivalent plot for the EWs of He II λ3835 + Hη and He II λ4860 + Hβ. This provides a rough estimate of surface gravity, because of the decrease in line strength relative to the observed pseudo-continuum for short-wavelength Balmer lines as they broaden as a result of increasing log(*g*). The observed sample is consistent with surface gravities log(*g*) ≳ 5, higher than is observed in MS stars.

Figure 3C shows the pure helium blend He I + He II λ4026 as a function of the hydrogen/helium blend He II + Hδ λ4100, which probe the hydrogen and helium surface mass fractions. The observed stars are all consistent with hydrogen-depleted surfaces, spanning the location of the model grid from $X_{\rm H,surf}$ = 0.01 (almost hydrogen-free) to $X_{\rm H,surf}$ = 0.3. We chose these diagnostics to avoid more wind-sensitive spectral lines. Our tests with different assumptions for the mass-loss rate and wind velocity do not change these results (*20*).

The properties we estimated for each star are listed in table S2. These diagnostics indicate that the class 1 stars are hot, compact, and hydrogen-poor. Figure 1 shows that their brightnesses fall along a sequence, connecting WR stars and the slightly lower luminosity WN3/O3 stars (*36*) to subdwarfs. Figure S13 shows that this sequence also appears in the strengths of stellar wind lines in the optical spectra.

Figure 4 compares our derived constraints on $T_{\rm eff}$ and log(*g*) with predictions for intermediate-mass helium stars (*22*). The observed stars have surface gravities between those of MS stars and white dwarfs—consistent with our expectations for helium stars—and temperatures hotter than most subdwarf stars (*37*). Figure 4 also shows a set of evolutionary tracks (*22*). The observed stars are consistent with predictions for the core-helium burning phase of ~2.5 to 8 $M_\odot$ stripped stars, which have progenitors with initial masses of between ~9 and 25 $M_\odot$. These ranges are high enough for the stars to later undergo core collapse (*38*), so they will explode as stripped-envelope supernovae (*39*). The winds from stars with initial masses of <25 $M_\odot$ are too weak to remove the hydrogen-rich envelope (*40*), so binary interaction is thought to be the primary mechanism for stripping stars in that mass range (see supplementary text section of the supplementary materials).

Excluding alternative explanations

We considered other possible interpretations of the stars in our sample (see supplementary text). Some of the class 3 stars could be ordinary (B-type) MS stars in regions with very little dust obscuration; in this case, they would appear to show a UV excess resulting from an overcorrection of the photometry. However, the same is not true for the class 1 and class 2 stars, whose locations on the CMD (Fig. 1) are inconsistent with massive O-type MS stars (the only type of MS star expected to have detectable He II) for any extinction values. The absorption-line spectral morphologies of the class 1 and class 2 stars are also distinct from those of WR stars (Fig. 2A).

Other types of stars can reach very high temperatures and similar brightnesses, such as accreting white dwarfs, central stars of planetary

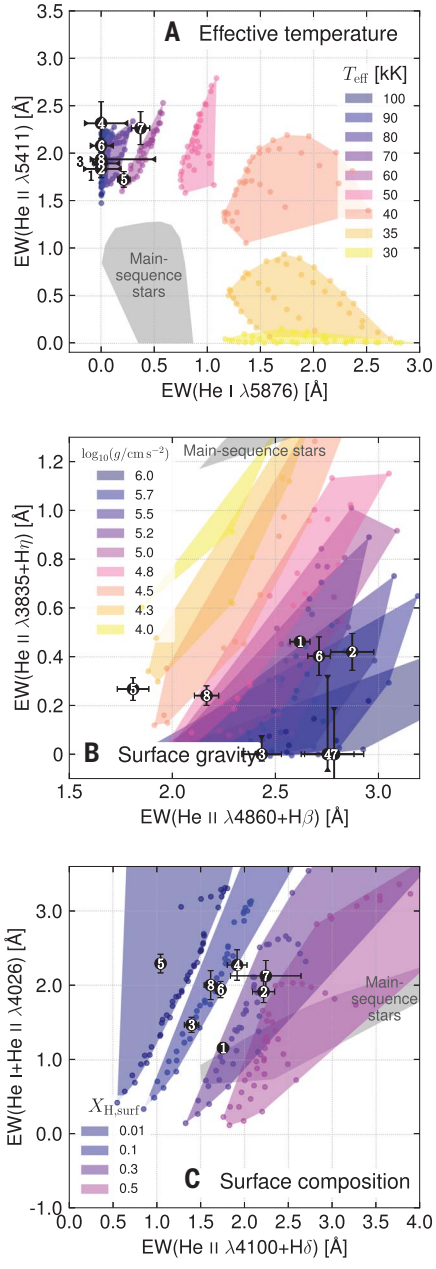


Fig. 3. Diagnostic diagrams used to estimate stellar properties. Each panel plots EWs of pairs of spectral lines (or line blends), chosen to constrain (**A**) effective temperature, (**B**) surface gravity, and (**C**) helium enrichment. Black data points (with 1σ error bars) show the measured EWs for the eight class 1 stars, which have helium star–type spectra. Colored dots show predictions from a grid of model spectra, with properties indicated by color (see legends; shaded regions encompass the full range for each value). Model equivalent widths were calculated assuming a signal-to-noise of 100, consistent with the median signal-to-noise of the observed class 1 stars (*20*). Gray shaded regions indicate MS star models (*20*). As shown in (A), all the class 1 stars have an effective temperature of >50 kK, so we only show models with $T_{\rm eff} \geq 50$ kK and O-type MS models in (B) and (C). The class 1 stars are hot, compact, hydrogen-poor, and do not overlap MS stars.

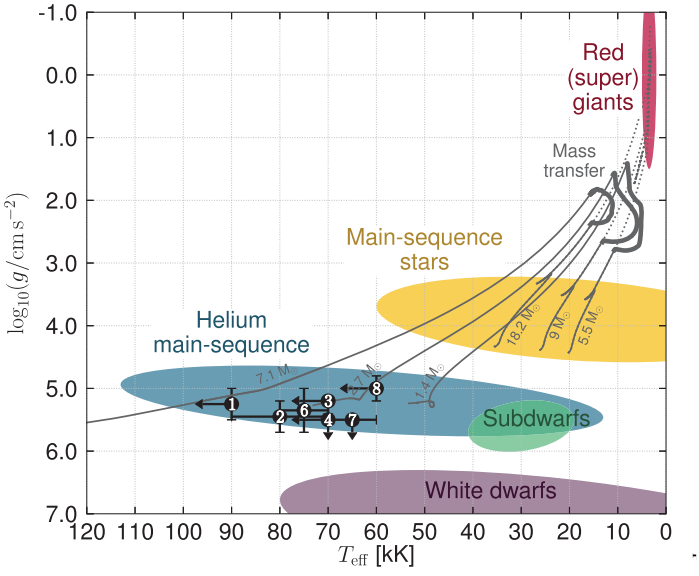


Fig. 4. Physical properties of the eight class 1 systems compared with other types of stars. Estimated effective temperatures and surface gravities of the class 1 stars are shown as numbered black circles. Error bars and limits correspond to the properties of the models that the star overlaps with in Fig. 3 and are listed in table S2. Colored ellipses indicate the regions occupied by main-sequence stars (yellow), white dwarfs (purple), theoretical predictions for helium stars fusing helium in the center (the helium main-sequence, blue), subdwarfs (green), and red giants and supergiants (red). Black lines show evolutionary tracks for stars with ZAMS masses of 5.5, 9.0, and 18.2 $M_\odot$, starting on the main sequence and extending until the end of central helium burning (*20*). Single stars of these masses (dotted lines) would evolve into cool and extended red giants or red supergiants. Binary stars of these initial masses undergo envelope stripping by means of mass transfer (thick solid lines) and subsequently evolve to burn helium as hot and compact helium stars with stripped masses of 1.4, 2.7, and 7.1 $M_\odot$ (thin solid lines). These stripped helium stars have surface properties similar to our observed sample.

nebulae, and very young post-asymptotic giant branch (post-AGB) stars (*10*), but those types all have circumstellar material, which produces emission lines or an infrared excess (*41*, *42*), neither of which we observe. Young post-AGB stars are also expected to be very rare (see supplementary text). Very fast rotation could fully mix stars—resulting in hot and compact helium stars—but this is only expected at higher masses and luminosities (*43*). Some hot low-mass objects (such as evolved subdwarfs and white dwarf merger products) could pollute our sample, but our targeting of the LMC and SMC means that they would need to be foreground objects located in the halo of the Milky Way. By examining the frequency of UV excesses in a control sample, we predict that there are <1 foreground objects along the line of sight to the LMC or SMC with colors, magnitudes, and kinematics similar to our spectroscopic sample (see supplementary text).

The properties we infer for the observed stars differ from previously identified helium-rich stars. They are much hotter and more compact than cool helium giants, such as the star ν Sgr [$T_{\rm eff} \lesssim 15$ kK, $\log(g) \sim 2$] (*44*, *45*). While the mass of ν Sgr is uncertain (*46*), it could be in a subsequent evolutionary phase; some stripped stars expand substantially upon completion of core-helium burning (*47*). The surface properties we estimated for the class 1 stars are similar to those of HD 45166, but their spectra are distinct, with HD 45166 having a spectrum dominated by emission lines. This indicates that the anomalously slow wind speed observed in HD 45166 might not be common (see supplementary text). Instead, the absorption spectra of stars in our sample imply low mass-loss rates, consistent with theoretical predictions (*20*).

The properties, binary companions, and evolutionary history of the individual systems in our sample are likely diverse. Nevertheless, we conclude that they constitute a population of massive stars stripped through binary interaction. Because only a subset of stripped star binaries are expected to show a UV excess (*20*), the population we observe represents only a small fraction of the predicted intermediate-mass helium stars. Many other examples could be hidden by brighter companion stars. With estimated masses of ~2 to 8 $M_\odot$, the stars we observed fill a gap in previously identified helium stars, connecting subdwarfs with WR stars.

REFERENCES AND NOTES

1. H. Sana *et al.*, *Science* **337**, 444–446 (2012).
2. M. Moe, R. Di Stefano, *Astrophys. J. Suppl. Ser.* **230**, 15 (2017).
3. Y. Shao, X.-D. Li, *Astrophys. J.* **908**, 67 (2021).
4. J. D. Lyman *et al.*, *Mon. Not. R. Astron. Soc.* **457**, 328–350 (2016).
5. N. Smith, W. Li, A. V. Filippenko, R. Chornock, *Mon. Not. R. Astron. Soc.* **412**, 1522–1538 (2011).
6. J. J. Eldridge, M. Fraser, S. J. Smartt, J. R. Maund, R. M. Crockett, *Mon. Not. R. Astron. Soc.* **436**, 774–795 (2013).
7. T. M. Tauris *et al.*, *Astrophys. J.* **846**, 170 (2017).
8. E. R. Stanway, J. J. Eldridge, G. D. Becker, *Mon. Not. R. Astron. Soc.* **456**, 485–499 (2016).
9. Y. Götberg *et al.*, *Astron. Astrophys.* **634**, A134 (2020).
10. C. S. Jeffery, in *Hydrogen-Deficient Stars*, A. Werner, T. Rauch, Eds., vol. 391 of *Astronomical Society of the Pacific Conference Series* (Astronomical Society of the Pacific, 2008), pp. 3–16.
11. P. A. Crowther, *Annu. Rev. Astron. Astrophys.* **45**, 177–219 (2007).
12. U. Heber, *Publ. Astron. Soc. Pac.* **128**, 082001 (2016).

13. C. S. Jeffery, in *Hydrogen-Deficient Stars*, A. Werner, T. Rauch, Eds., vol. 391 of *Astronomical Society of the Pacific Conference Series* (Astronomical Society of the Pacific, 2008), pp. 53–63.
14. W. A. Weidmann, R. Gamen, *Astron. Astrophys.* **526**, A6 (2011).
15. L. Wang *et al*., *Astron. J.* **161**, 248 (2021).
16. M. M. Shara *et al*., *Mon. Not. R. Astron. Soc.* **464**, 2066–2074 (2017).
17. M. H. van Kerkwijk *et al*., *Nature* **355**, 703–705 (1992).
18. J. H. Groh, A. S. Oliveira, J. E. Steiner, *Astron. Astrophys.* **485**, 245–256 (2008).
19. T. Shenar *et al*., *Science* **381**, 761–765 (2023).
20. Materials and methods are available as supplementary materials.
21. P. Massey, K. F. Neugent, N. Morrell, D. J. Hillier, *Astrophys. J.* **788**, 83 (2014).
22. Y. Götberg *et al*., *Astron. Astrophys.* **615**, A78 (2018).
23. L. M. Z. Hagen *et al*., *Mon. Not. R. Astron. Soc.* **466**, 4540–4557 (2017).
24. D. Lang, D. W. Hogg, D. J. Schlegel, *Astron. J.* **151**, 36 (2016).
25. D. Zaritsky, J. Harris, I. B. Thompson, E. K. Grebel, P. Massey, *Astron. J.* **123**, 855–872 (2002).
26. D. Zaritsky, J. Harris, I. B. Thompson, E. K. Grebel, *Astron. J.* **128**, 1606–1614 (2004).
27. K. D. Gordon, G. C. Clayton, K. A. Misselt, A. U. Landolt, M. J. Wolff, *Astrophys. J.* **594**, 279–293 (2003).
28. J. L. Marshall *et al*., *Proc. SPIE* **7014**, 701454 (2008).
29. Gaia Collaboration, *Astron. Astrophys.* **674**, A1 (2023).
30. R. O. Gray, C. J. Corbally, *Stellar Spectral Classification*, Princeton Series in Astrophysics (Princeton Univ. Press, 2009).
31. J. S. Vink, *Astron. Astrophys.* **607**, L8 (2017).
32. D. J. Hillier, D. L. Miller, *Astrophys. J.* **496**, 407–427 (1998).
33. T. Lanz, I. Hubeny, *Astrophys. J. Suppl. Ser.* **146**, 417–441 (2003).
34. T. Lanz, I. Hubeny, *Astrophys. J. Suppl. Ser.* **169**, 83–104 (2007).
35. N. R. Walborn *et al*., *Astron. J.* **123**, 2754–2771 (2002).
36. K. F. Neugent, P. Massey, D. J. Hillier, N. Morrell, *Astrophys. J.* **841**, 20 (2017).
37. J. S. Drilling, C. S. Jeffery, U. Heber, S. Moehler, R. Napiwotzki, *Astron. Astrophys.* **551**, A31 (2013).
38. T. Ertl, S. E. Woosley, T. Sukhbold, H. T. Janka, *Astrophys. J.* **890**, 51 (2020).
39. S.-C. Yoon, L. Dessart, A. Clocchiatti, *Astrophys. J.* **840**, 10 (2017).
40. E. R. Beasor, B. Davies, N. Smith, *Astrophys. J.* **922**, 55 (2021).
41. H. Van Winckel, *Annu. Rev. Astron. Astrophys.* **41**, 391–427 (2003).
42. W. R. Hamann, M. Peña, G. Gräfener, M. T. Ruiz, *Astron. Astrophys.* **409**, 969–982 (2003).
43. I. Brott *et al*., *Astron. Astrophys.* **530**, A115 (2011).
44. D. Schoenberner, J. S. Drilling, *Astrophys. J.* **268**, 225–227 (1983).
45. T. Kipper, V. G. Klochkova, *Balt. Astron.* **21**, 219–240 (2012).
46. A. Gilkis, T. Shenar, *Mon. Not. R. Astron. Soc.* **518**, 3541–3555 (2023).
47. E. Laplace, Y. Götberg, S. E. de Mink, S. Justham, R. Farmer, *Astron. Astrophys.* **637**, A6 (2020).
48. J. Breysacher, M. Azzopardi, G. Testor, *Astron. Astrophys. Suppl. Ser.* **137**, 117–145 (1999).
49. P. Massey, A. S. Duffy, *Astrophys. J.* **550**, 713–723 (2001).
50. W.-R. Hamann, L. Koesterke, U. Wessolowski, *Astron. Astrophys. Suppl. Ser.* **113**, 459–471 (1995).
51. M. R. Drout *et al*., An observed population of intermediate-mass helium stars that have been stripped in binaries – theoretical, computational and observational data, Zenodo (2023); https://doi.org/10.5281/zenodo.7600727.
52. Y. Götberg *et al*., Stellar properties of observed stars stripped in binaries in the Magellanic Clouds – spectral models, Zenodo (2023); https://doi.org/10.5281/zenodo.7976200.
53. M. R. Drout *et al*., An observed population of intermediate-mass helium stars that have been stripped in binaries – raw spectroscopic data, Zenodo (2023); https://doi.org/10.5281/zenodo.10035849.

ACKNOWLEDGMENTS

We thank K. Auchettl, K. Breivik, C. Burns, A. Carpenter, J. Fuller, B. Hovis-Afflerbach, A. Ji, C. Johnston, D. Kelson, M. van Kerkwijk, D. Lang, C. Norman, T. Piro, M. Renzo, A. Roc, P. Senchyna, S. Torres, and M. Zapartas for fruitful discussions and support, and we acknowledge feedback from the anonymous referees that improved this manuscript. This paper includes data gathered with the 6.5 m Magellan Telescopes located at Las Campanas Observatory, Chile. We thank J. Mulchaey (Carnegie Observatories director), L. Infante (Las Campanas Observatory director), and the entire Las Campanas staff for their hard work and dedication to keeping the observatory operating through the COVID-19 pandemic, when a large fraction of this data was obtained. Computing resources used for this work were made possible by a grant from the Ahmanson Foundation. **Funding:** M.R.D. acknowledges support from the NSERC through grant RGPIN-2019-06186, the Canada Research Chairs Program, the Canadian Institute for Advanced Research (CIFAR), and the Dunlap Institute at the University of Toronto. Y.G. was supported by NASA through the NASA Hubble Fellowship Program grant HST-HF2-51457.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. A.J.G.O. acknowledges support from the Lachlan Gilchrist Fellowship Fund. S.E.d.M. acknowledges funding by the Netherlands Organization for Scientific Research (NWO) as part of the Vidi research program BinWaves under project number 639.042.728. N.S. was supported by NASA through HST grant GO-15824. **Author contributions:** M.R.D., S.E.d.M., and Y.G. designed the UV-optical color-magnitude diagrams used for candidate identification. M.R.D. and B.A.L. performed photometry on the Swift UVOT images and selected candidate targets for follow-up. M.R.D., Y.G., B.A.L., and A.J.G.O. performed the spectroscopic observations with the Magellan Baade telescope. M.R.D. reduced and analyzed the spectroscopic data and performed the kinematic analysis. Y.G. computed the spectral models with CMFGEN with input from J.H.G. and designed the equivalent width diagnostics used to estimate stellar properties. M.R.D. and Y.G. interpreted the observed sample in the context of the stellar models. M.R.D. and Y.G. wrote the paper, with input from B.A.L., J.H.G., S.E.d.M., A.J.G.O., and N.S. **Competing interests:** The authors have no competing interests to declare. **Data and materials availability:** Reduced and combined spectra of the 25 stars in our spectroscopic sample, our stripped helium star models, our main-sequence spectral models, equivalent width measurements from the theoretical models, and the index files used for astrometry are archived at Zenodo (*51*, *52*). The UV and optical photometry for the observed stars and models are provided in tables S6, S4, and S7 and are archived in machine-readable form at Zenodo (*51*). Raw Magellan Echellette spectra are also archived at Zenodo (*53*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade4970
Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 to S9
References (*54*–*167*)

Submitted 30 August 2022; accepted 27 October 2023
10.1126/science.ade4970

THERMOREGULATION

Self-sustaining personal all-day thermoregulatory clothing using only sunlight

Ziyuan Wang[1]†, Yiwen Bo[2]†, Peijia Bai[2], Shuchao Zhang[1], Guanghui Li[1], Xiangjian Wan[1], Yongsheng Liu[1]*, Rujun Ma[2]*, Yongsheng Chen[1]*

The human body must stay within a certain temperature range for comfort and safety. However, challenges for thermoregulatory clothing exist for harsh application scenarios, such as full day/night cycles, frigid polar regions, and space travel. We developed a flexible and sustainable personal thermoregulatory clothing system by integrating a flexible organic photovoltaic (OPV) module to directly acquire energy from sunlight and bidirectional electrocaloric (EC) devices. The flexible OPV-EC thermoregulatory clothing (OETC) can extend the human thermal comfort zone from 22°–28°C to 12.5°–37.6°C with a fast thermoregulation rate. The low energy consumption and high efficiency of the EC device allows for 24 hours of controllable and dual-mode thermoregulation with 12 hours of sunlight energy input. This self-powered wearable thermoregulatory platform has a simple structure, compact design, high efficiency, and strong self-adaptability with sunlight as the sole energy source.

Clothing plays an indispensable role in regulating body heat for preserving the thermal comfort of our body in daily life (*1*). One of the most common scenarios is to maintain body temperature within a safe range in situations of fluctuating and sometimes rapid changes in environmental temperature, for instance, walking from a comfortable indoor environment (~25°C) to a hot (>36°C) or cold (<15°C) outdoor environment. Without the ability to quickly adapt to such fast-changing environmental temperatures and cool down or warm up, people may become uncomfortable or sick and may even die (*2*). The more challenging scenarios are to keep our bodies in a comfortable temperature range (skin temperature) in harsh environments, such as frigid polar regions or space travel (which is extremely hot in sunlight but extremely cold in the dark). Therefore, wearable thermoregulatory clothing, capable of keeping the human body in a comfortable temperature range (skin temperature), as a spacesuit does, has been a long-sought but challenging goal for smart clothing systems.

Indeed, many thermoregulatory systems have been developed, which can be broadly classified into passive and active systems. Passive thermoregulatory systems include radiative thermoregulatory systems (*3*–*8*), phase-change thermoregulatory systems (*9*, *10*), and adsorption thermoregulatory systems (*11*, *12*). However, most systems powered by solar energy with self-sustainability can achieve only one-way thermoregulation (*3*–*5*, *8*). The systems that have bidirectional thermoregulation need to improve their efficiencies, response speeds and tunable temperature ranges (skin temperature) (*6*, *7*, *13*–*15*).

Active thermoregulatory systems allow for rapid cooling or warming of the human body. In general, cooling vests are based on coolant circulation or water/ice fluidic channels that allow for wearable thermoregulation; however, these systems require large and complex mechanical compressors (*16*, *17*). Although some excellent solid-state active thermoregulatory systems have been developed and the need for compressors and traditional liquid or vapor refrigerants could be eliminated, they still bear some substantial limits. For example, although the Joule effect heater is effective in heating with a controlled temperature, the high power consumption and lack of cooling capability strongly limit its application (*18*–*20*). The thermoregulatory systems based on magnetocaloric (*21*, *22*) and elastocaloric effects (*23*, *24*) require a large magnetic field and high mechanical load to realize good thermal-management effect, respectively, and thus have limited wearability. Thermoelectric (TE) thermoregulatory devices (*25*–*28*) based on the Peltier effect have found various applications. However, these devices generally exhibit low efficiency because of their high energy consumption. In particular, they all need extra energy input and cannot be kept working for a long period of time without additional energy. Although battery-based thermoregulatory systems can achieve good thermal-management performance in a short period of time, their limited energy supply cannot provide all-day and sustained thermoregulation for the human body (*25*, *27*).

Therefore, developing an all-day, self-powered, bidirectional thermoregulatory clothing system capable of rapidly responding to various complex or quick environmental temperature changes and keeping the human body in a comfortable temperature zone remains a challenging goal (*29*). The systems must harvest energy to achieve all-day thermoregulation. One of the more obvious sources of sustainable energy source is sunlight, which can be collected anywhere in open space. Indeed, solar cells, including an organic photovoltaic (OPV) cell, have evolved substantially (*30*), and the state-of-the-art OPV cell with a size of 4 mm^2 has a power conversion efficiency (PCE) of >20% (*31*). Simultaneously, flexible and high-performance OPV cells with PCEs >17% (*32*) have been demonstrated, which can be integrated into clothing to collect solar energy (*33*, *34*).

To achieve the required sustainability and flexibility as well as light weight, the thermal-management unit for the body must be highly efficient in transferring energy and have a low energy consumption. Therefore, we selected recently developed electrocaloric (EC) devices, which have high efficiency, low energy consumption, and bidirectional thermoregulatory properties and are pollution free (*35*–*38*). For example, one flexible EC thermoregulatory device has a very low energy cost, could reach a coefficient of performance of 13, and has a specific cooling power of 2.8 W/g (*35*).

We choose a flexible OPV module powered by sunlight and a high-efficiency heat transfer EC device as the two main units to fabricate a self-sustained thermoregulatory clothing system. We aim to power it only by solar energy with the capability of all-day (24 hours) cycling between hot/light and cold/dark environments. Our flexible OPV-EC thermoregulatory clothing (OETC) system exhibits highly efficient and fast performance in both cooling and warming modes as needed. Moreover, it can extend the thermal comfort zone by 19.1 K (from 6.0 to 25.1 K) and reach intelligent and controllable all-day dual-mode thermoregulation for the human body as needed. With the combination of these features, the human body wearing our OETC system can quickly adapt, as needed, to the changes of environmental temperature during outdoor activities and even possibly in such as harsh environments as polar regions or personal space travel.

Results and discussion

We fabricated a large flexible OPV module (*39*, *40*) with a thickness of only 180 μm for the sunlight energy–collecting unit in our OETC system (*41*). The entire flexible OPV module with an effective area of 25.2 cm^2 could provide a total voltage of 5.75 V and PCE of 11.85% under standard air mass 1.5 global (AM 1.5G, 100 mW/cm^2) (figs. S1 to S3 and table S1).

For the thermoregulatory unit in our OETC, we selected poly(vinylidene fluoride-trifluoroethylene-chlorofluoroethylene) [P(VDF-TrFE-CFE)], mainly because of its large entropy change, large

[1]State Key Laboratory and Institute of Elemento-Organic Chemistry, The Centre of Nanoscale Science and Technology and Key Laboratory of Functional Polymer Materials, Renewable Energy Conversion and Storage Center (RECAST), College of Chemistry, Nankai University, Tianjin 300071, China. [2]School of Materials Science and Engineering, National Institute for Advanced Materials, Nankai University, Tianjin 300350, China.
*Corresponding author. Email: yschen99@nankai.edu.cn (Y.C.), malab@nankai.edu.cn (R.M.), liuys@nankai.edu.cn (Y.L.)
†These authors contributed equally to this work.

adiabatic temperature change near room temperature, and good mechanical flexibility (*42*). The EC system based on P(VDF-TrFE-CFE) has substantial potential for efficient thermoregulation (*35*, *36*, *43*). We fabricated a flexible EC thermoregulatory device following Ma *et al.* (*35*, *44*) (fig. S4). Noticeably, our flexible EC device exhibits the same thermal-management performance as the rigid one (figs. S5 and S6).

With these two flexible units ready, we integrated them together for the OETC system (fig. S7). In sunlight, the OPV module efficiently converts solar energy into electrical energy to drive the EC device directly to provide a cooling effect (Fig. 1). The excess energy can be stored in a simple attached energy storage system (ESS) (fig. S8) because of the low energy consumption of the EC device, as discussed below. The power provided by the OPV is sufficient to power the entire OETC system (see fig. S8 for details). Therefore, in the dark, our OETC system can use the stored energy provided by the ESS to maintain body temperature when the environment is cold and thus achieve a full day (day/night) of operation. The cooling mode and warming mode can be switched as desired at any time to achieve individual thermal comfort.

Working mechanism of the OETC system in cooling/warming mode

We show a photograph of the flexible OETC thermoregulation system assembled by one OPV module and two EC units (Fig. 2A). This compact assembly mode can provide effective cooling/warming for the human body as needed. The working mechanism of the OETC system for the cooling mode (Fig. 2B) is the same as that powered by the electric supply (*37*), but in our system, we power it directly by the electricity generated by the OPV module (see fig. S9 for details). The cooling mode includes the following steps (*35*): (i) electrostatic actuation of the EC polymer stack toward the top flexible heat transfer layer (as a heat sink with large heat capacity) (fig. S10); (ii) the EC polymer stack is heated up by applying an electric field on the EC polymer stack, and thus, heat transfers from the EC polymer stack to the flexible heat transfer layer [Fig. 2B, (1)]; (iii) electrostatic actuation of the EC polymer stack toward the bottom human skin (as a heat source); (iv) the EC polymer stack is cooled down by removing the electric field, and thus, heat transfers from the human skin to the EC polymer stack to realize one cycle of skin cooling [Fig. 2B, (2)]. For the warming mode, warming is achieved by changing the heat transfer to the opposite direction by changing the sequence of the four steps described above, which is realized by simply adjusting the phase of square-wave voltage. Correspondingly, the warming mode has similar steps to the cooling mode but with the opposite heat transfer effect: (i) electrostatic actuation of the EC polymer stack toward the bottom human skin that needs to be warmed up; (ii) the EC polymer stack is heated up by applying an electric field on the EC polymer stack, and thus, heat transfers from the EC polymer stack to the human skin (as a heat sink) [Fig. 2B, (3)]; (iii) electrostatic actuation of the EC polymer stack toward the top flexible heat transfer layer (as a heat source); (iv) the EC polymer stack is cooled down by removing the electric field, and thus, heat transfers from the flexible heat transfer layer to the EC polymer stack to finish one cycle of skin warming [Fig. 2B, (4)]. With these two working modes, bidirectional controllable thermoregulation for cooling and warming can be implemented as needed.

Electrostatic actuation is a simple and fast method to control the heat transport speed by adjusting the working frequency of the EC device (*35*–*37*, *43*). We compared the temperature span of the OETC system at different frequencies under the standard AM 1.5G (100 mW/cm^2) by a solar simulator (fig. S11). Although the OETC system can operate at higher frequencies, the frequency of the OETC system that gives a maximum temperature span of 2.9 K is 0.75 Hz (one complete cycle takes ~1.33 s), in part because of the time needed to transfer the heat from EC stack to human skin and flexible heat transfer layer.

Temperature span of the OETC system in different working scenarios

The temperature span of our OETC thermoregulation can also be easily adjusted by the illumination intensity. With the increase of illumination intensity, the flexible OPV module can reach higher voltage (power), and thus, the input voltage of the EC device increases (fig. S12), which results in a higher thermoregulatory performance of the OETC system. We measured the temperature difference (ΔT, difference between real-time temperature and initial temperature) of the OETC system under different illumination intensities of 55 (fig. S13), 70, and 100 mW/cm^2 irradiation using a solar simulator at a frequency of 0.75 Hz (Fig. 2C). The OETC system works well at different illumination intensities, and the maximum temperature span can reach 2.9 K when the illumination intensity is standard AM 1.5G sunlight (100 mW/cm^2). Furthermore, the outdoor thermoregulation performance of our OETC system was also demonstrated by direct solar radiation under clear sky conditions from 9:00 to 16:00 in Tianjin, China (3 August 2022) (fig. S14). Although the intensity of outdoor sunlight varies considerably with time, our OETC system still shows good and stable thermal-management ability at different illumination intensities. When the outdoor sunlight intensity is the same as the simulated illumination intensity, the OETC system exhibits almost the same thermal management (Fig. 2C and fig. S14B). The whole process runs without external energy sources and realizes self-powered thermoregulation with zero energy consumption.

Although an outside electric supply was previously required to power the EC device to achieve the effective thermal management reported in the literature (*35*–*37*, *43*), we demonstrated that the EC device can instead be powered directly on site by an integrated flexible OPV module. The integrated device shows the same outstanding performance, including the same temperature difference (ΔT) at the same electric field (Fig. 2D).

Excellent sustainable performance of the OETC system

We compared the thermoregulatory performance of the commercial rigid TE device of the same size as the EC device powered by a

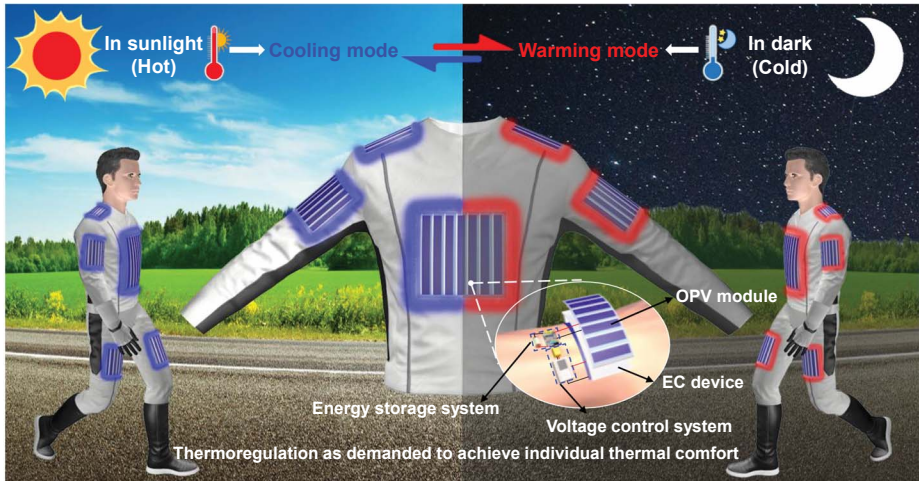


Fig. 1. Working schema when wearing our flexible OETC to achieve individual thermal comfort in a cycle between hot (in sunlight) and cold (in dark) environments as demanded.

flexible OPV module under an illumination intensity of 100 mW/cm^2 (Fig. 2E and fig. S15). The temperature span (Fig. 2E) and heat flux (figs. S15C and S16) of our OETC system are 2.9 K and 28.76 mW/cm^2, respectively, whereas the OPV-TE system shows only the temperature span of 1.2 K (Fig. 2E and fig. S15B) and heat flux of 16.79 mW/cm^2 (fig. S15C), respectively, under the same illumination intensity of 100 mW/cm^2. In addition, we also compared the thermoregulatory performance of the perovskite photovoltaic module with a similar size to that of the EC device powered by an OPV module under the illumination intensity of 100 mW/cm^2 (fig. S17). Compared with the thermal-management performance of the EC device powered by OPV module in Fig. 2C, the EC device powered by the perovskite solar module (fig. S17) displayed almost same results. Meanwhile, we calculated the power consumption of the EC device under different illumination intensities (fig. S18). Under the illumination intensity of 100 mW/cm^2, the average power consumption of the EC device is only 1.91 mW/cm^2 at 0.75 Hz because of its low energy consumption. Considering that the PCE of our OPV module (with an active area of 25.2 cm^2) is 11.85% under standard AM 1.5G (100 mW/cm^2) and the energy consumption of the EC device (with an active area of 8 cm^2) is only 15.28 mW (1.91 mW/cm^2 × 8 cm^2 = 15.28 mW), a simple estimate indicates that the total generated electricity is 298.58 mW (100 mW/cm^2 × 11.85% × 25.2 cm^2 = 298.58 mW). Thus, we benefit from the low energy consumption (15.28 mW) of the EC device, with 283.30 mW (298.58 mW − 15.28 mW = 283.30 mW) of surplus energy that could be stored under ideal conditions (figs. S8 and S19). The surplus energy stored in the ESS could be automatically switched to power the entire system at night with no extra energy input to realize a full day/night thermoregulatory cycle (fig. S20). Moreover, it is worthwhile to note that energy recovery is also possible during the depolarization process of the EC effect (*43*), which thus further improves the efficiency of our OETC system.

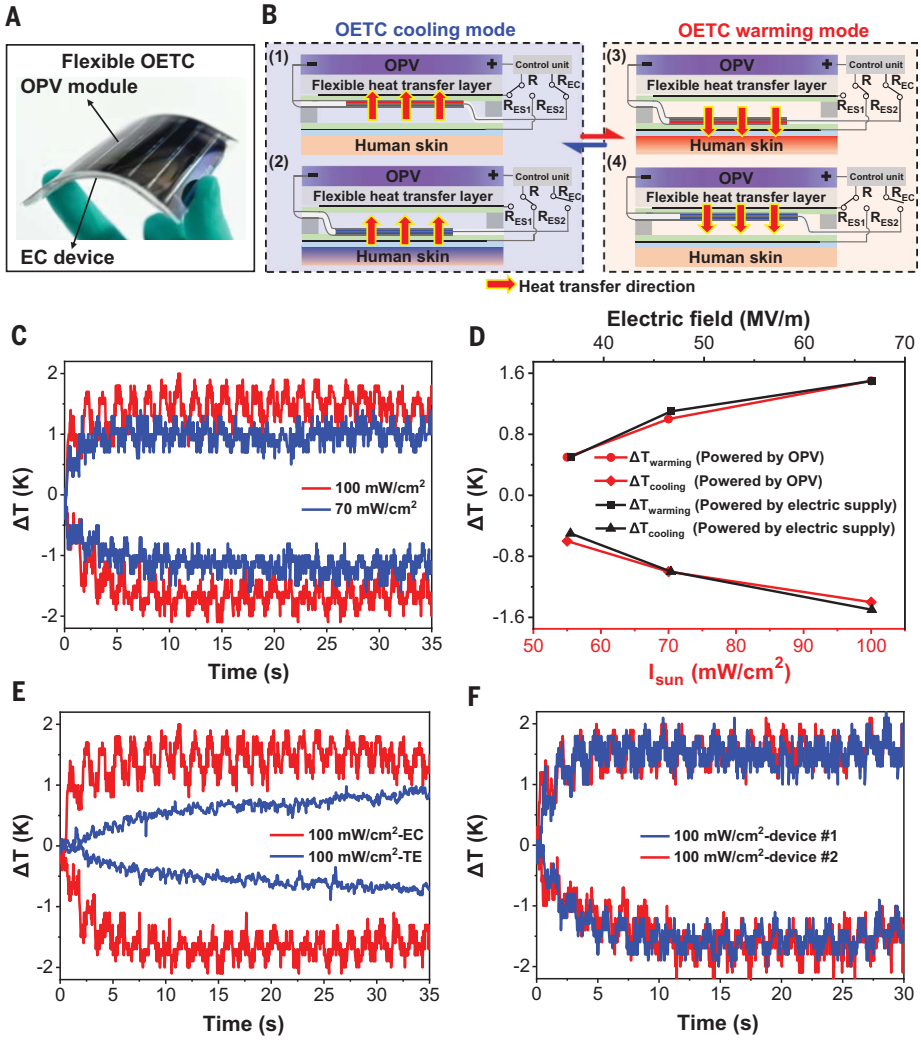


Fig. 2. Performance of the flexible OETC system. (**A**) Photograph of an OETC system assembled by an OPV module and two EC units. (**B**) Working mechanism of the OETC system in cooling/warming mode, respectively. The cooling/warming mode can be easily switched by the control unit when moving from a thermal comfort environment into a hot/cold environment. R_{ES}, the relay for controlling the electrostatic actuation. (**C**) Temperature span of the OETC system under different illumination intensities (70 and 100 mW/cm^2) in cooling/warming mode at 0.75 Hz. (**D**) Comparison of ΔT of the EC thermoregulation device driven by OPV module and electric supply under different electric fields. (**E**) Comparison of temperature span of EC and TE thermoregulation devices with the same size (active area of 8 cm^2) driven by the same OPV module under an illumination intensity of 100 mW/cm^2. (**F**) Temperature span of two EC parallel array devices (active area of 16 cm^2) driven simultaneously by one OPV module (active area of 25.2 cm^2) under an illumination intensity of 100 mW/cm^2.

Performance of an OPV-EC array

The EC devices have good array cooperativity, and one single OPV module with an active area of 25.2 cm^2 has sufficient power to simultaneously drive two parallel arrays of EC devices with an active area of 16 cm^2. For example, under standard AM 1.5G (100 mW/cm^2), the two EC parallel devices could be synchronized completely, and both could reach a temperature span of 2.9 K, which demonstrates their bidirectional thermoregulatory performance (Fig. 2F). To further extend its application in wearable thermoregulation, we evaluated the performance of four parallel EC arrays driven by one OPV module under an illumination intensity of 100 mW/cm^2 (fig. S21). Four parallel EC arrays can simultaneously achieve bidirectional controllable thermoregulation (movie S1). This indicates that our OETC system has good scalability required for the practical wearable thermoregulation.

Thermoregulatory performance of OETC on the human body

To demonstrate the wearability of the OETC for meeting the flexible needs of human body thermoregulation, we measured the stability of OETC cooling and warming mode performance in the bending state (Fig. 3A and fig. S22). During the bending measurement, the system is illuminated with a 100 mW/cm^2 light at the bottom, and the surface temperature is measured by an infrared camera at the top (movie S2). Our OETC reaches its maximum and stable thermoregulation performance when it starts to operate at 0.75 Hz for 10 s. The initial state of the OETC is flat, and the radius of curvature (k) of OETC is 0 m^{-1}. Then, the OETC system is bent at a uniform rate of 0.12 m^{-1} s^{-1} to reach a maximum curvature of 3.6 m^{-1}, followed by the same rate of bending release until the curvature of the OETC returns to 0 m^{-1}. During the operation, we observed negligible change in its thermoregulatory performance in the flat, bent, and released states of the OETC, which demonstrates excellent flexibility.

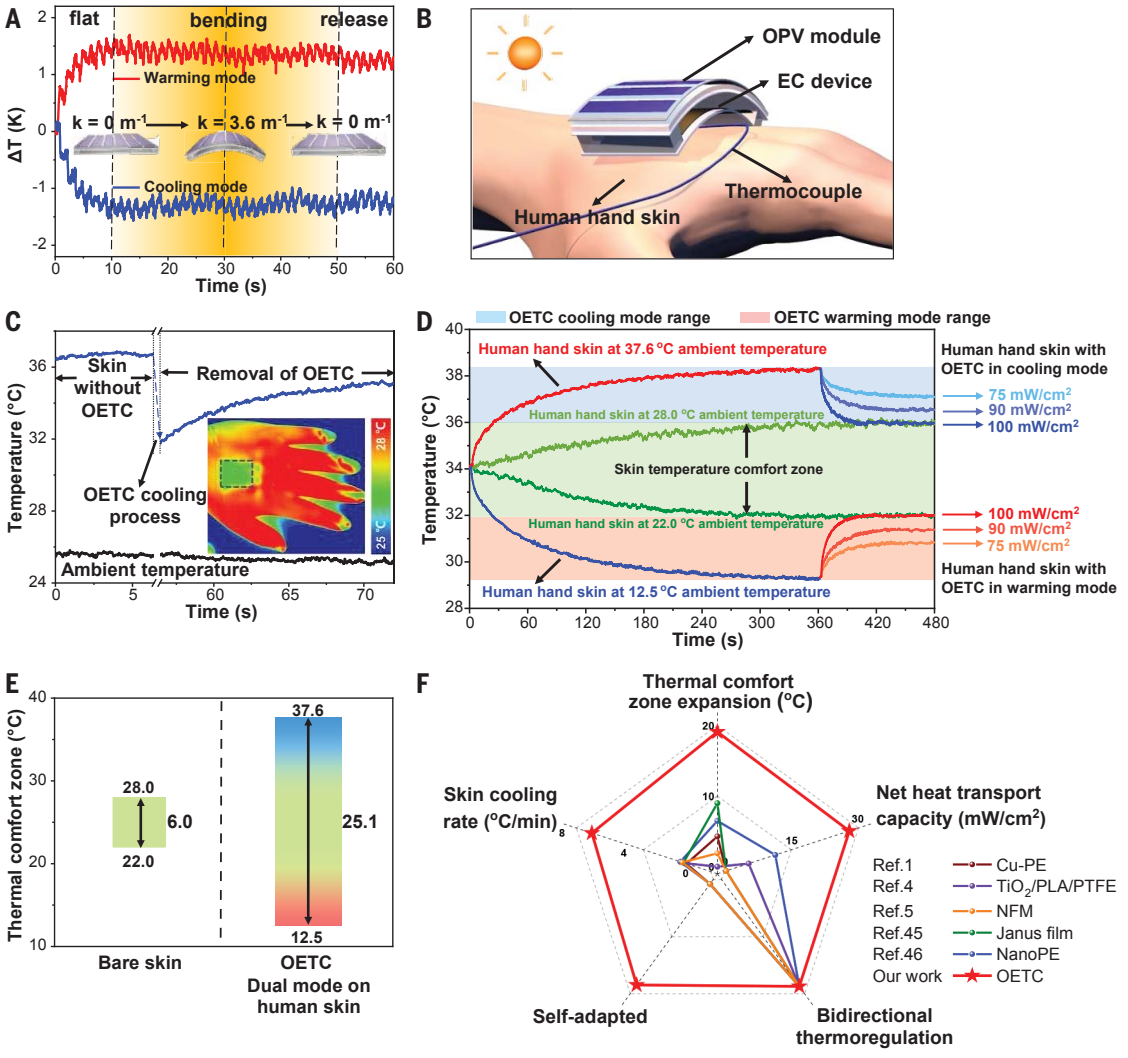


Fig. 3. Wearable thermoregulatory performance of OETC. (**A**) Performance of the flexible OETC under different curvature. (**B**) Schematic of the thermoregulatory setup consists of the skin, a thermocouple measuring the temperature of the skin, and OETC covering the skin. (**C**) Thermoregulation of human hand in OETC cooling mode. The inset is the infrared thermal image of a human hand in OETC cooling mode. (**D**) Thermoregulation of the human hand skin by OETC under different sunlight intensity (75, 90, 100 mW/cm^2) at different environmental temperatures. The initial temperature of the skin is 34.0°C. When moving the skin into different environmental temperatures of 37.6° and 12.5°C, the skin can reach 38.3° and 29.2°C, respectively. By using the OETC cooling mode at high environmental temperature and warming mode at low temperature under standard AM 1.5G (100 mW/cm^2), the human skin temperature can be maintained in the thermal comfort temperature range of 32.0° to 36.0°C (skin temperature), even though the environmental temperature changes between 12.5° and 37.6°C. When the illumination intensity is lower than 100 mW/cm^2 (75 or 90 mW/cm^2), our OETC system still has bidirectional thermoregulation performance. (**E**) Thermal comfort zone of bare human skin and human skin with OETC dual working mode. (**F**) Comparison of the net power, thermal comfort zone expansion, cooling capability for skin, bidirectional thermoregulation, and self-adaptability of the OETC with related representative works reported in the literature (*1*, *4*, *5*, *45*, *46*). The definitions of Cu-PE, TiO_2/PLA/PTFE, NFM, Janus film, and NanoPE are given in the supplementary text.

We further applied the flexible OETC to human skin for thermoregulation. We show the experimental setup of the flexible OETC thermal measurement on the human skin (Fig. 3B) and the thermoregulation of a human hand in OETC cooling mode (Fig. 3C). We monitored the whole process with an infrared camera detector under an illumination intensity of 100 mW/cm^2 at an environmental temperature of 26°C. Our OETC cooled the human skin from 36.8° to 31.7°C at an average rate of 6.1°C/min to achieve fast thermoregulation (Fig. 3C).

The human body must remain within a certain temperature range (skin temperature) for a comfortable and safe existence, but this range varies individually (*45*–*47*). We set a comfort range on the basis of observed human skin temperatures between 32° and 36°C (*46*), which requires an environmental temperature range between 22° and 28°C (thus, the thermal comfort zone of bare human skin is 6.0 K in our measurement) (Fig. 3D).

We measured the thermoregulation performance on human skin directly, with the human hand temperature starting at 34.0°C and a corresponding environmental temperature of 25.0°C (the middle point of the comfort zone). Under standard AM 1.5G (100 mW/cm^2), when moving the skin into a low-temperature environment (12.5°C), the skin temperature drops to 29.2°C and the OETC warming mode starts working, which raises the skin temperature up to the thermal comfort temperature of 32.0°C (Fig. 3D). Correspondingly, when moving the skin into a higher-temperature environment of 37.6°C, the skin temperature rises to 38.3°C. The OETC cooling mode turns on, which brings the skin temperature down to the thermal comfort temperature of 36.0°C.

As a result, our OETC maintains human skin temperature within a thermal comfort zone between 32.0°C and 36.0°C, even though the environmental temperature varies between 12.5° and 37.6°C. Compared with bare human skin (thermal comfort zone of 6 K), our OETC extends the thermal comfort zone of the skin by 19.1 K (Fig. 3E) for this module size and illumination intensity. In addition, skin can be warmed at a maximum rate of 15.6°C/min or cooled at a maximum rate of 14.0°C/min in the first 5 s to achieve fast thermoregulation. When the illumination intensity is lower than 100 mW/cm^2 (75 or 90 mW/cm^2), our OETC system still has bidirectional thermoregulation performance. Under an illumination intensity of 90 mW/cm^2, the OETC warming

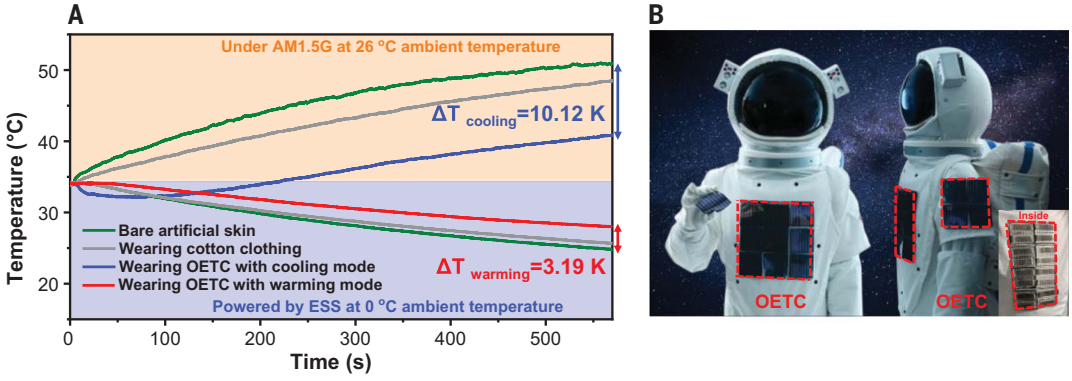


Fig. 4. The thermoregulation performance of OETC compared with cotton clothing and prospect of personal space travel. (**A**) Temperature changes of bare artificial skin, skin wearing cotton clothing, and skin wearing OETC in the sunlight of 100 mW/cm^2 at a 26°C environment temperature and in the dark at a 0°C environment temperature, respectively. The initial temperature of the artificial skin is 34°C, and the temperature is measured by thermocouples. (**B**) Photograph of the OETC worn on the human body for the prospect of personal space travel.

mode can raise the skin temperature from 29.4° to 31.3°C, whereas the OETC cooling mode could lower the skin temperature from 38.3° to 36.5°C. These temperatures fall only slightly outside the thermal comfort zone. When the illumination intensity is 75 mW/cm^2, the OETC system still can bidirectionally thermoregulate the skin from 29.4° to 30.8°C in warming mode and from 38.3° to 37.2°C in cooling mode. Furthermore, we also evaluate the thermoregulation performance of OETC on artificial skin at 100 mW/cm^2 under different environmental temperatures (fig. S23). The thermal comfort zone of bare artificial skin is from 23° to 27°C (4.0 K) (*46*); OETC extends the thermal comfort zone of the artificial skin by 16.6 K (fig. S23A). Although OETC cannot restore the temperature of artificial skin to the thermal comfort zone in a harsher environment, it still has good thermoregulation performance (fig. S23B). Improvements in the thermal comfort zone can be made with performance or efficiency improvements in the OPV or EC units. Alternatively, the relative ratios and sizes of the OPV and EC units could be further optimized.

We summarize the net heat transport capacity, thermal comfort zone expansion, skin cooling rate, bidirectional thermoregulation, and self-adaptability of our OETC compared with related representative works reported in the literature (Fig. 3F) (*1*, *4*, *5*, *45*, *46*). By benefiting from the efficient net heat transport capacity of 27.89 mW/cm^2 (fig. S16), our OETC can extend the human thermal comfort zone by 19.1 K. Moreover, the OETC system can effectively cool the human skin at an average rate of 6.1°C/min to achieve fast thermoregulation. By benefiting from the low energy consumption of the EC device, the OETC system could operate a full 24 hours with 12 hours of sunlight energy input (fig. S18). Thus, the combined characteristics of our personal thermoregution system, such as controllable, all-day dual mode and notable thermoregulation performance, could make it possible for individuals to stay more comfortable in harsh environments while using only sunlight as the energy source.

Thermoregulation performance of OETC in the outdoors and the prospect for use in space

We measured and compared the temperature changes of bare artificial skin, skin covered with cotton clothing, and skin covered with an OETC (Fig. 4A) in sunlight of 100 mW/cm^2 at 26.0°C environmental temperature and in the dark at 0°C environmental temperature, respectively. Under an illumination intensity of standard AM 1.5G (100 mW/cm^2) at 26.0°C environmental temperature, the temperature of bare skin and skin covered with cotton clothing can be raised from 34.0°C up to 50.9° and 48.4°C, respectively. Nevertheless, the temperature of the artificial skin covered with an OETC is only 40.8°C. The maximum cooling capacity reaches 10.1 K (calculated by the temperature difference between bare artificial skin and artificial skin covered with an OETC in cooling mode after 570 s exposed to an illumination intensity of 100 mW/cm^2 in a 26.0°C environmental temperature), which demonstrates the cooling power of OETC. In addition, OETC can be driven to warm the skin by using the ESS in cold night at a 0°C environmental temperature. The warming performance of artificial skin covered with an OETC compared with artificial skin is 3.2 K (calculated by the temperature difference between bare artificial skin and artificial skin covered with an OETC in warming mode after 570 s exposed to the 0°C environmental temperature) higher than those of skin covered with cotton clothing and bare skin, which demonstrates the excellent warming capacity of OETC.

The bidirectional thermoregulation that uses solar power could make this device of interest for integrating into a conventional spacesuit to help reduce the overall power requirements (Fig. 4B). During individual space travel, the theoretical area of a spacesuit is around 1.85 m^2 (*48*). In space, the magnitude of the solar radiation pressure depends on the solar flux near the surface of Earth, and a solar constant of 136.7 mW/cm^2 is normally used to calculate the solar flux in 1 astronomical unit (*49*). With continued improvements in solar cell performance, including that of the flexible OPV module, if we assume that a 45% PCE solar cell device is used, we estimate that an OPV module to provide all-day human body thermoregulation will have an area of only 1.12 m^2 (*50*). We believe that this OETC system could be optimized in the future in terms of both performance and practicability for application in harsher environments. The temperature span of the EC device can be increased to improve the thermoregulation performance of our OETC system. First, for the material side, the double bond–modified P(VDF-TrFE-CFE) materials could provide a larger temperature change of 7.8 K at 118 MV/m (*37*). Second, the device could be optimized by using a cascade device to increase the temperature span of 4.8 K (double-deck) and 8.7 K (four-layer cascade) (*38*, *45*). Lastly, by adding nanofillers to improve the thermal conductivity of P(VDF-TrFE-CFE) (*51*) or by using an active EC regenerator to further increase the temperature span (*52*, *53*), the EC performance can be further improved. Clearly, further studies are needed to develop a practical product based on the prototype and concept demonstrated in this work.

Conclusions

We developed an advanced self-powered wearable thermoregulatory system that integrates flexible OPV module and EC thermoregulation units together for efficient personalized thermoregulation. Its active control feature can be used for fast cooling/warming dual-mode thermoregulation as needed by the human body. Moreover, the thermal comfort zone can be extended from 6.0 to 25.1 K by OETC with rapid thermoregulation, which can ensure the safety and comfort of the human body in various complex and unstable environments. By benefiting from the low energy consumption of the EC device, OETC can achieve controllable and all-day dual-mode thermoregulation. Together with its other outstanding features such as simple and compact structure, high efficiency, and strong self-adaptability, with more optimization, we believe that the OETC could

demonstrate potential applications in the field of high-end thermoregulation and even extend human survivability in harsh environments such as polar regions and individual space walking.

REFERENCES AND NOTES

1. J. Chai *et al*., *Cell Rep. Phys. Sci.* **3**, 100958 (2022).
2. ANSI/ASHRAE, *Standard 55: Thermal Environmental Conditions for Human Occupancy* (American Society of Heating, Refrigerating, and Air-Conditioning Engineers, 2017).
3. X. A. Zhang *et al*., *Science* **363**, 619–623 (2019).
4. S. Zeng *et al*., *Science* **373**, 692–696 (2021).
5. R. Xiao *et al*., *ACS Appl. Mater. Interfaces* **11**, 44673–44681 (2019).
6. Q. Zhang *et al*., *Nat. Commun.* **13**, 4874 (2022).
7. Q. Zhang, Y. Wang, Y. Lv, S. Yu, R. Ma, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2207353119 (2022).
8. P.-C. Hsu *et al*., *Science* **353**, 1019–1023 (2016).
9. Y. Kou *et al*., *Energy Storage Mater.* **34**, 508–514 (2021).
10. Z. Luo *et al*., *Adv. Funct. Mater.* **33**, 2212032 (2023).
11. C. Wang *et al*., *Joule* **4**, 435–447 (2020).
12. J. A. Mason *et al*., *Nature* **527**, 357–361 (2015).
13. L. Zhou *et al*., *Cell Rep. Phys. Sci.* **2**, 100338 (2021).
14. H. Luo *et al*., *Nano Lett.* **21**, 3879–3886 (2021).
15. L. Zhang *et al*., *Angew. Chem. Int. Ed.* **54**, 3952–3956 (2015).
16. M. Mokhtari Yazdi, M. Sheikhzadeh, *J. Text. Inst.* **105**, 1231–1250 (2014).
17. M. Smith, V. Cacucciolo, H. Shea, *Science* **379**, 1327–1332 (2023).
18. G.-X. Li, T. Dong, L. Zhu, T. Cui, S. Chen, *Chem. Eng. J.* **453**, 139763 (2023).
19. Y. Guo *et al*., *Small* **13**, 1702645 (2017).
20. Z. Guo, C. Sun, J. Wang, Z. Cai, F. Ge, *ACS Appl. Mater. Interfaces* **13**, 8851–8862 (2021).
21. D. J. Silva, B. D. Bordalo, A. M. Pereira, J. Ventura, J. P. Araújo, *Appl. Energy* **93**, 570–574 (2012).
22. Y. Liu *et al*., *Nat. Commun.* **7**, 11614 (2016).
23. S. Zhang *et al*., *Nat. Commun.* **13**, 9 (2022).
24. R. Wang *et al*., *Science* **366**, 216–221 (2019).
25. J. Choi, C. Dun, C. Forsythe, M. P. Gordon, J. J. Urban, *J. Mater. Chem. A Mater. Energy Sustain.* **9**, 15696–15703 (2021).
26. H. Wei, J. Zhang, Y. Han, D. Xu, *Appl. Energy* **326**, 119941 (2022).
27. Y. Zhang *et al*., *ACS Appl. Mater. Interfaces* **14**, 15317–15323 (2022).
28. R. A. Kishore, A. Nozariasbmarz, B. Poudel, M. Sanghadasa, S. Priya, *Nat. Commun.* **10**, 1765 (2019).
29. R. Mutschler, M. Rüdisüli, P. Heer, S. Eggimann, *Appl. Energy* **288**, 116636 (2021).
30. L. Meng *et al*., *Science* **361**, 1094–1098 (2018).
31. J. Wang *et al*., *Natl. Sci. Rev.* **10**, nwad085 (2023).
32. G. Zeng *et al*., *J. Am. Chem. Soc.* **144**, 8658–8668 (2022).
33. H. Jinno *et al*., *Nat. Energy* **2**, 780–785 (2017).
34. A. Sharma *et al*., *Appl. Mater. Today* **29**, 101614 (2022).
35. R. Ma *et al*., *Science* **357**, 1130–1134 (2017).
36. Y. Bo *et al*., *Adv. Energy Mater.* **11**, 2003771 (2021).
37. P. Bai *et al*., *Adv. Mater.* **35**, e2209181 (2023).
38. H. Cui *et al*., *Joule* **6**, 258–268 (2022).
39. C.-Y. Liao *et al*., *Joule* **4**, 189–206 (2020).
40. S. Zhang *et al*., *Sol. RRL* **7**, 2300029 (2023).
41. See "Preparation of the flexible large-area OPV module" in the materials and methods for the detailed procedure.
42. B. Neese *et al*., *Science* **321**, 821–823 (2008).
43. Y. Meng *et al*., *Nat. Energy* **5**, 996–1002 (2020).
44. See "Preparation of the flexible EC device" in the materials and methods for the detailed procedure.
45. B. Dai, X. Li, T. Xu, X. Zhang, *ACS Appl. Mater. Interfaces* **14**, 18877–18883 (2022).
46. P.-C. Hsu *et al*., *Sci. Adv.* **3**, e1700895 (2017).
47. F. Salata, I. Golasi, V. Ciancio, F. Rosso, *Build. Environ.* **146**, 50–63 (2018).
48. J. G. Metts, J. A. Nabity, D. M. Klaus, *Adv. Space Res.* **47**, 1256–1264 (2011).
49. O. Montenbruck, P. Steigenberger, U. Hugentobler, *J. Geod.* **89**, 283–297 (2014).
50. See "Calculation of the required OPV module area to provide all-day thermoregulation for individual space travel" in the supplementary text for details.
51. G. Zhang *et al*., *Adv. Mater.* **27**, 1450–1454 (2015).
52. A. Torelló *et al*., *Science* **370**, 125–129 (2020).
53. Y. Wang *et al*., *Science* **370**, 129–133 (2020).

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support from the National Key R&D Program of China (2022YFB4200400, 2019YFA0705900, 2020YFA0711500), the National Natural Science Fund of China (21935007, 52025033, 51973095, and 52273248), and the Key Project of Natural Science Foundation of Tianjin City (21JCZDJC00010). **Funding:** National Key R&D Program of China (2022YFB4200400, 2019YFA0705900, 2020YFA0711500); National Natural Science Fund of China (21935007, 52025033, 51973095, and 52273248); and Key Project of Natural Science Foundation of Tianjin City (21JCZDJC00010). **Author contributions:** Y.C. and R.M. conceived and designed the project. Z.W. and Y.B. fabricated the OETC thermoregulation system and carried out the OETC performance studies. S.Z. and Z.W. fabricated the OPV module. Y.B. and P.B. fabricated the EC thermoregulation system. Z.W., Y.B., G.L., X.W., Y.L., R.M., and Y.C. analyzed and interpreted the data. The manuscript was mainly prepared by Y.C., R.M., Z.W., and Y.B. All authors reviewed and commented on the manuscript. **Competing interests:** Y.C., R.M., Y.L., Z.W., Y.B., and Y.Z. are inventors on a China provisional patent application serial number 2023110782432, related to this work. The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj3654
Materials and Methods
Supplementary Text
Figs. S1 to S23
Table S1
References (*54*–*59*)
Movies S1 and S2

Submitted 23 June 2023; accepted 24 October 2023
10.1126/science.adj3654

MACHINE LEARNING

Backpropagation-free training of deep physical neural networks

Ali Momeni[1], Babak Rahmani[2], Matthieu Malléjac[1], Philipp del Hougne[3], Romain Fleury[1]*

Recent successes in deep learning for vision and natural language processing are attributed to larger models but come with energy consumption and scalability issues. Current training of digital deep-learning models primarily relies on backpropagation that is unsuitable for physical implementation. In this work, we propose a simple deep neural network architecture augmented by a physical local learning (PhyLL) algorithm, which enables supervised and unsupervised training of deep physical neural networks without detailed knowledge of the nonlinear physical layer's properties. We trained diverse wave-based physical neural networks in vowel and image classification experiments, showcasing the universality of our approach. Our method shows advantages over other hardware-aware training schemes by improving training speed, enhancing robustness, and reducing power consumption by eliminating the need for system modeling and thus decreasing digital computation.

Deep learning has emerged as a breakthrough technology with outstanding success (*1*, *2*) that primarily operates on traditional von Neumann computing hardware. This technology is currently facing high energy consumption, such as the 1.3–gigawatt-hour (GWh) electricity usage of GPT-3 (*3*), and low computing speed (*4*). Because of these challenges, researchers are exploring alternative physical platforms for artificial neural networks (ANNs), including optics (*5*–*9*), spintronics, (*10*, *11*), nanoelectronic devices (*12*–*15*), photonic hardware (*5*), and acoustic systems (*16*, *17*).

Two primary methods currently dominate neural network hardware design. The first involves designing hardware to implement trained mathematical transformations through strict operation-by-operation mathematical isomorphism, primarily targeting the inference phase of deep learning (*18*–*21*). The second category, deep physical neural networks (PNNs), focuses on training the hardware's physical transformations directly to perform the desired computations. PNNs hold the promise of more scalable, energy-efficient, and faster neural network hardware by exploiting physical transformations and eliminating the conventional software-hardware separation (*22*, *23*).

So far, the training of PNNs has predominantly relied on backpropagation (BP) (*24*). Yet, there are several reasons why BP is not a suitable choice for PNNs, one of which is the complexity and lack of scalability in the physical implementations of BP operations in the hardware (*25*–*28*). Commonly, PNN proposals use in silico training, performing BP calculations on an external computer with a digital twin of the physical system. However, this method may result in potential simulation-reality gaps as a result of inaccurate representation of the physical system (*6*–*8*, *10*, *13*, *14*, *20*, *29*, *30*). Moreover, physics-aware training methods based on BP (PA-BP) (*22*) offer improvements over traditional in silico methods but still necessitate a differentiable digital model for the backward pass. Additionally, PA-BP–trained PNNs may face challenges when subjected to strong perturbations, potentially rendering fine-tuned models unusable and necessitating retraining from scratch.

Another important drawback of BP is its reliance on having complete knowledge of the computation graph carried out during the forward pass to accurately compute derivatives (*23*, *31*–*34*). When a black box is inserted in the forward pass, BP becomes infeasible. Therefore, alternative training methods for PNNs have proved advantageous. For example, an approach explored for training physical networks is the augmented direct feedback alignment (DFA) method (*23*), which aims to avoid the need for a differentiable digital model. However, this method is only compatible with certain physical networks, where it is possible to separate the nonlinear and linear layers.

Local learning has been extensively studied for training digital neural networks, from early work on Hebbian contrastive learning in Hopfield models (*35*) to recent biologically plausible frameworks (*31*, *34*, *36*, *37*), blockwise BP (*38*, *39*), and contrastive representation learning (*40*, *41*). Inspired by this concept and to address the limitations of BP-based PNN training, we proposed a simple and physics-compatible PNN architecture augmented by a physical local learning (PhyLL) algorithm. The proposed method enables supervised and unsupervised contrastive learning training of arbitrary PNNs locally without the need to know the nonlinear physical layers and train a digital twin model. In this BP-free method, the standard backward pass, typically performed by a digital computer, is replaced with an additional single forward pass through a physical system. This substitution can improve training speed, power consumption, and memory usage during the training phase of wave-based PNNs by eliminating the extra overhead incurred because of the digital twin modeling phase present in other hardware-aware frameworks. We showed the robustness and adaptability of the proposed method, even in systems exposed to unpredictable external perturbations. To showcase the universality of our approach, we performed experimental vowel and image classification using three wave-based systems that differ in terms of the underlying wave phenomenon and the type of nonlinearity involved (a detailed description of each system can be found in the supplementary text, sections 2.3 to 2.5).

PhyLL

Figure 1A shows a simple and physics-compatible deep PNN including N nonlinear physical data transformers augmented by trainable linear multiplications. Each nonlinear physical data transformer performs a nonlinear mapping between the input and output followed by an augmented trainable linear multiplication to classify distinct classes through a local training algorithm. The output of each layer is then passed to the next layer. The subsequent layer then carries out the same process hierarchically on the output of the previous layer. The proposed architecture shares some similarities with conventional deep reservoir computing systems (*42*); see supplementary text, section 2.9, for further details on their differences.

The training algorithm is inspired by the recently proposed forward-forward algorithm (*31*) and local training proposals (*38*–*41*) in digital neural networks, which has been extended and adapted to the supervised and unsupervised model-free physical learning of PNNs. Each nonlinear physical system performs a nonlinear transformation on input data (Fig. 1), which can be expressed as $h^{(l)} = f_{\mathrm{N}}^{(l)}\left[W_{\mathrm{p}}^{(l)} x^{(l)}\right]$, where $x^{(l)}$, $W_{\mathrm{p}}^{(l)}$, and $f_{\mathrm{N}}^{(l)}$ correspond to the physical inputs (e.g., optical intensity, electric voltage, and vibration), the physical interconnections (e.g., optical, electrical, or mechanical coupling) in the physical system, and the physical nonlinearity (e.g., nonlinear optical, magnetic, or mechanical effects) in layer l, respectively.

Here, $W_{\mathrm{p}}^{(l)}$ and $f_{\mathrm{N}}^{(l)}$ denote the mixing operation and nonlinear kernel of the l–th physical

[1]Laboratory of Wave Engineering, Department of Electrical Engineering, EPFL, Lausanne CH-1015, Switzerland. [2]Microsoft Research, Cambridge CB4 0AB, UK. [3]University of Rennes, CNRS, IETR - UMR 6164, F-35000 Rennes, France.
*Corresponding author. Email: romain.fleury@epfl.ch

systems, respectively (supplementary text, sections 2.6 and 2.7). The output of layer l can be expressed as the multiplication of $h^{(l)}$ by the augmented trainable weight matrix $W_t^{(l)}$—i.e., $y^{(l)} = W_t^{(l)} h^{(l)}$. Such trainable matrix multiplications can be performed either digitally or through physical systems, for instance using Mach-Zehnder interferometer (MZI)–integrated photonics (*43*) or spatial light modulators (SLMs) in optics (*21*, *44*). The goal is to train $W_t^{(l)}$ locally without the need to know the nonlinear physical layer. Instead of a forward and backward pass, we use two physical forward passes: a positive and a negative forward pass through the physical system, each running on different physical inputs. The positive physical pass, $y_{pos}^{(l)} = W_t^{(l)} f_N^{(l)}\left[W_p^{(l)} x_{pos}^{(l)}\right]$, uses positive inputs that include the input dataset and the correct labels, and the negative physical pass, $y_{neg}^{(l)} = W_t^{(l)} f_N^{(l)}\left[W_p^{(l)} x_{neg}^{(l)}\right]$, uses negative inputs that include the input dataset and the incorrect labels (Fig. 1A). In each layer, we calculate the so-called goodness function, defined as the cosine similarity between the positive and negative activities. Eventually, for each layer l, $W_t^{(l)}$ is trained by minimizing the following loss function

$$L^{(l)} = \log(1 + \exp\{\theta[\cos_{sim}(y_{pos}, y_{neg})]\}) \quad (1)$$

In supervised learning, the goodness function is defined as the cosine similarity between the activities of the layer and a random vector drawn from normal distribution, both for the positive and negative physical passes. In this case, the loss function reads

$$L^{(l)} = \log\left[1 + \exp\left(-\theta\left\{\cos_{sim}\left[y_{pos}, \xi^{(l)}\right] - \cos_{sim}\left[y_{neg}, \xi^{(l)}\right]\right\}\right)\right] \quad (2)$$

In the equations above, $\cos_{sim}$ is the cosine similarity defined as the cosine of the angle between the two arguments, θ is a scale factor, and ξ^l is the random vector for the layer l. The original forward-forward algorithm uses only the difference of the positive and negative squared activities, hence necessitating layer normalization to be applied to the data before proceeding to subsequent layers (*31*). Conversely, our algorithm avoids incorporating layer

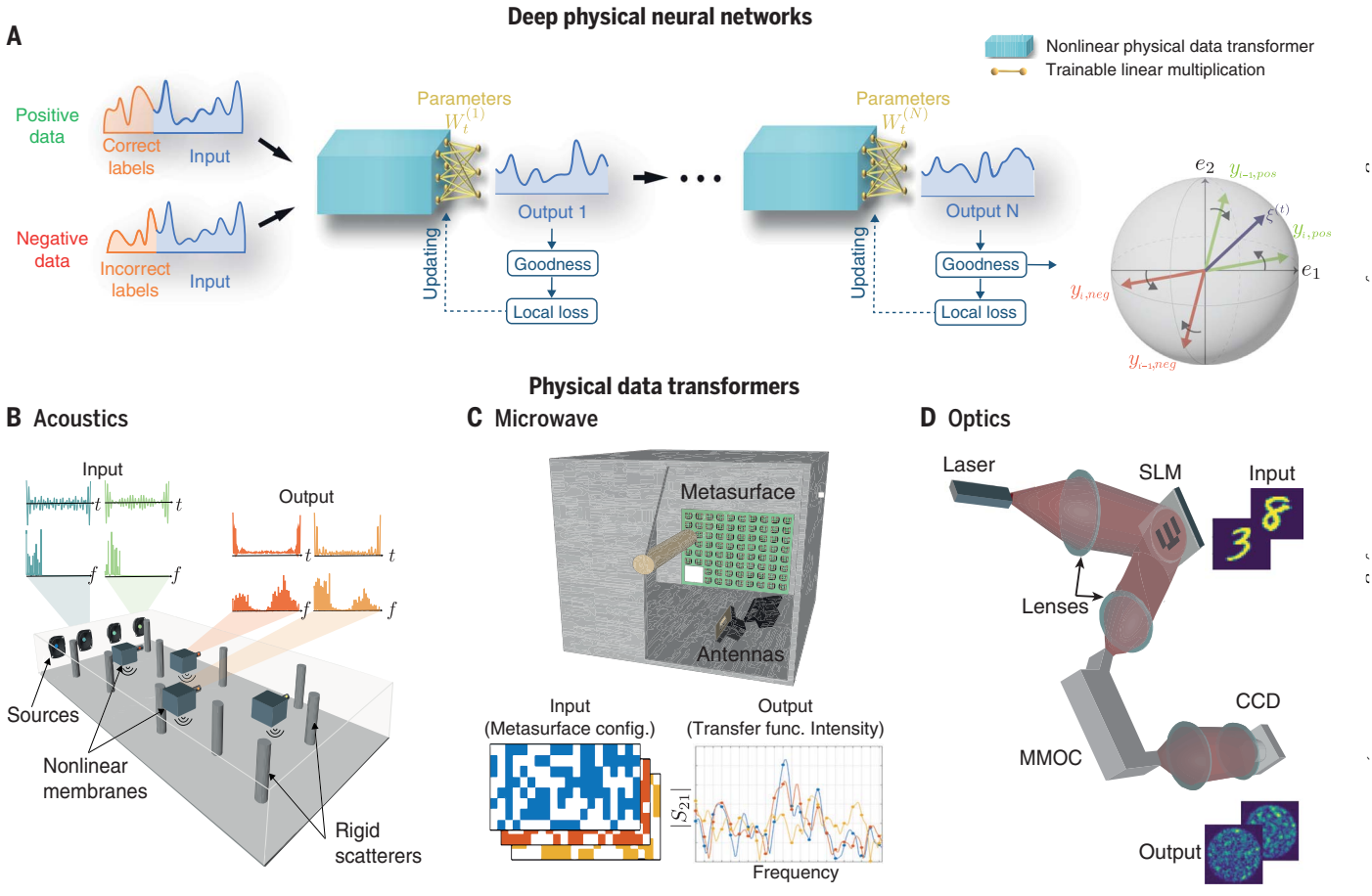


Fig. 1. Deep PNNs. (**A**) A simple and physics-compatible deep neural network that uses a sequence of nonlinear physical data transformers augmented by trainable matrix multiplications, trained by the supervised PhyLL technique (refer to supplementary text, section 2.1.1, for additional explanations). At each layer, the nonlinear physical data transformer conducts nonlinear mapping between input and output spaces to separate positive and negative data by maximizing the cosine similarity of the positive data to a random vector ξ and minimizing the cosine similarity of the negative data to the same vector. We considered three physical systems that vary in terms of the underlying wave phenomenon and the type of nonlinearity. (**B**) In acoustics, input data are encoded into the intensity of sound waves at different frequencies injected on the left side of the cavity. Sound waves propagate through a chaotic cavity that comprises multiple rigid cylindrical diffusers and nonlinear membranes. The transformed waveforms are received by multiple microphones. (**C**) In the chaotic microwave cavity, input data are encoded into the programmable metasurface configuration inside the metallic disordered cavity. The outputs are obtained from the waves' spectra (transfer function). (**D**) In the optical setup, input data are encoded onto the SLM, and after passing through a multimodal optical cavity (MMOC), the resulting optical intensity is measured on the charge-coupled device (CCD) camera [numerical experiment based on experimentally acquired data from Rahmani *et al.* (*56*)].

normalization into the architecture because its hardware-based implementation is a deep challenge. During the inference phase, we input a particular label into the PNNs and accumulate the goodness values for all layers. This process is repeated for each label separately. The label with the highest accumulated goodness value is then selected as the output (see supplementary text, section 2.1, for more details). In unsupervised contrastive learning, a single linear layer maps representations from pretrained hidden layers to labels (supplementary text, section 2.1.1). For details on data generation, refer to supplementary text, section 2.2.

Diverse PNNs for vowel and image classification

Figure 1 presents three deep PNN classifiers for various standard datasets, including vowel,

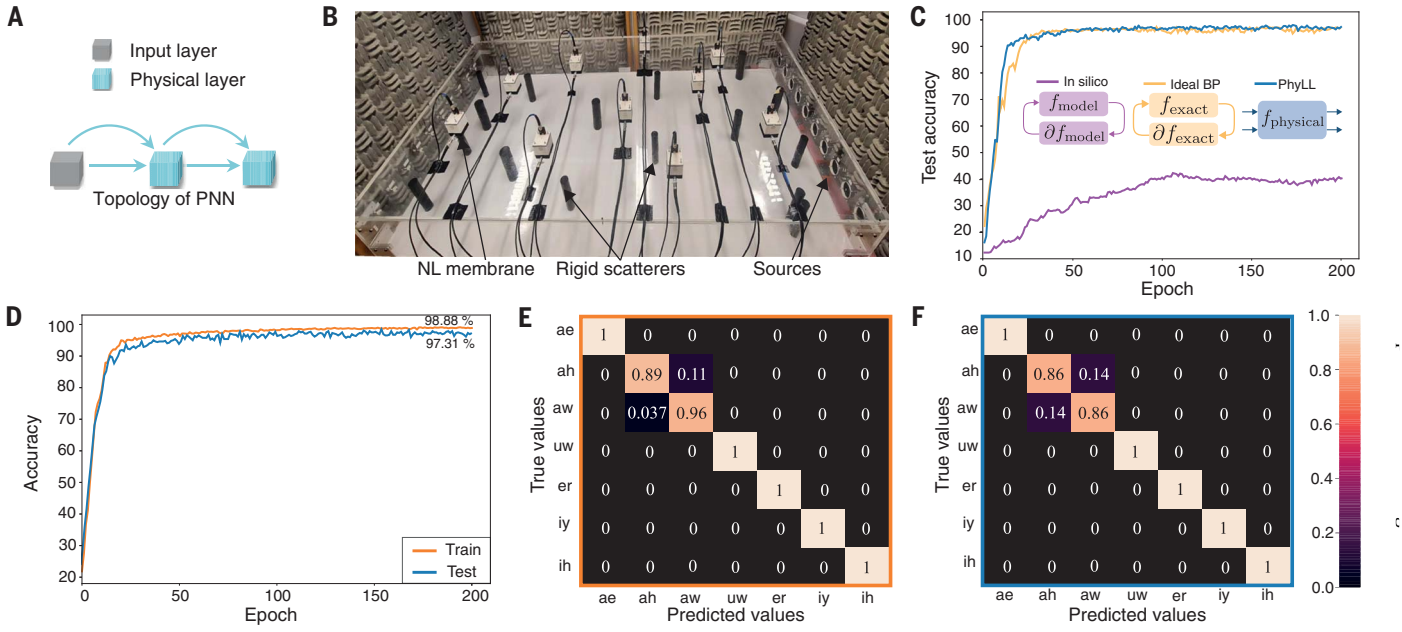


Fig. 2. Acoustic-PNN. (**A**) The topology of the acoustic-PNN consists of a two-layer PNN with skip connections. Each layer comprises an acoustic-PNN augmented by trainable matrix multiplication. (**B**) Photograph of the experimental setup. NL, nonlinear. (**C**) Comparison of test accuracy versus training epoch with in silico, ideal BP, and PhyLL algorithm for the vowel recognition task. (**D**) The train and test classification accuracy versus training epoch for the vowel recognition task. (**E** and **F**) The confusion matrix for the PNN on the train (E) and test (F) sets.

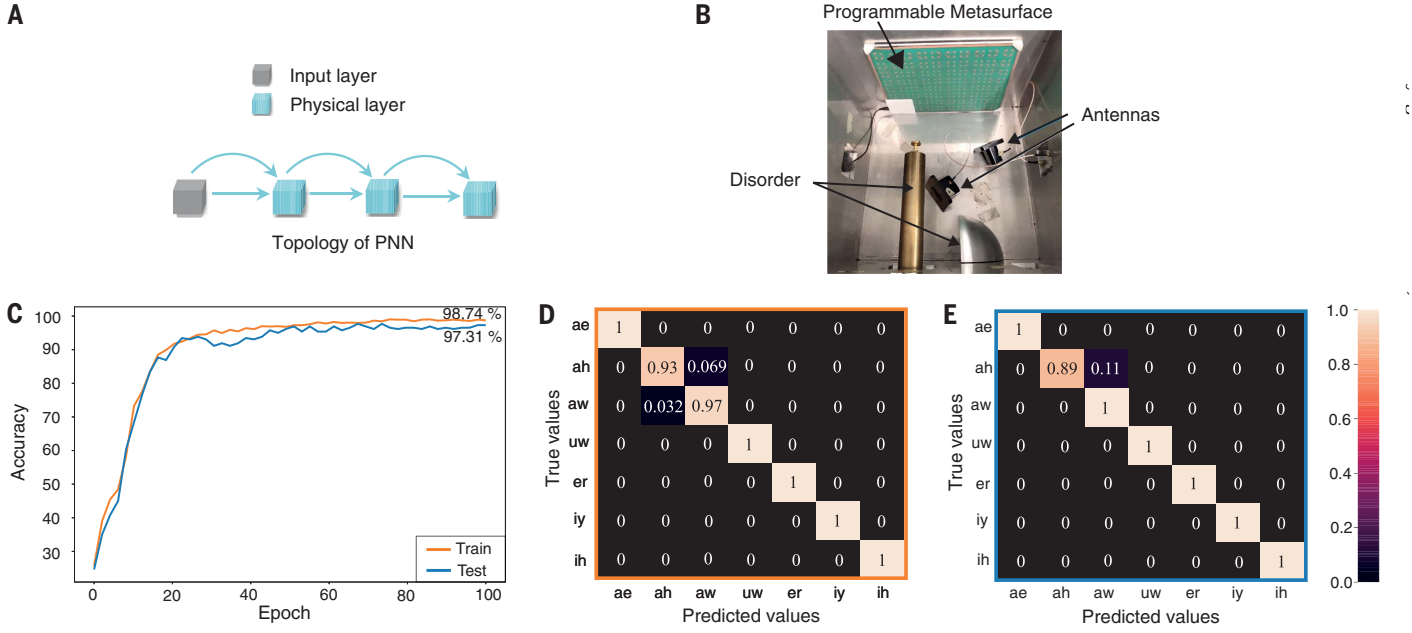


Fig. 3. Microwave-PNN. (**A**) The topology of the microwave-PNN consists of a three-layer PNN with skip connections. Each layer comprises a microwave-PNN augmented by trainable matrix multiplication. (**B**) Photograph of the experimental setup. (**C**) Train and test classification accuracy versus training epoch for the vowel recognition task. (**D** and **E**) Confusion matrix for the PNN on the train (D) and test (E) sets.

digit, fashion Mnist, and CIFAR10, based on three distinct physical systems, each featuring a distinctive source of nonlinearity (materials and methods and supplementary text, sections 2.3 to 2.5). Although there have been proposals that explore wave-based analog computing for linear operations, such as multiplication and convolution (*43*, *45*–*53*), it is important to note that PNNs require nonlinearity to effectively handle regression and classification tasks. We evaluated the performance of PhyLL in these media (refer to tables S1 and S2 in the supplementary materials) against in silico and BP methods under both supervised and unsupervised contrastive training schemes using an end-to-end surrogate forward model of the systems for benchmarking purposes.

Acoustic chaotic cavity with nonlinear scatterers

In acoustics, an air-filled multimode cavity composed of multiple nonlinear meta-scatterers randomly placed on the cavity top wall and multiple rigid scatterers inside the cavity was used (materials and methods and supplementary text, section 2.3). The nonlinear meta-scatterers were designed to provide a nonlinear relation between pressure and particle velocity with controllable power law. The positive and negative data were encoded onto the amplitude of each frequency component composing the excitation waveforms, which were then injected into the nonlinear system through loudspeakers positioned on the right side of the cavity. The output of the physical system was measured using microphones below the meta-scatterers. We investigated the vowel classification performance of two-layer acoustic-PNN (Fig. 2A). To compare the results of PhyLL with ideal BP and in silico training, we accurately modeled the forward pass of acoustic-PPN by a digital neural network (supplementary text, section 2.1). When trained using PhyLL, the acoustic-PNN achieved a classification accuracy of 98.88% and 97.31% for train and test datasets, respectively (Fig. 2, D to F). Figure 2C shows the comparison of the classification results obtained for PhyLL, ideal BP,

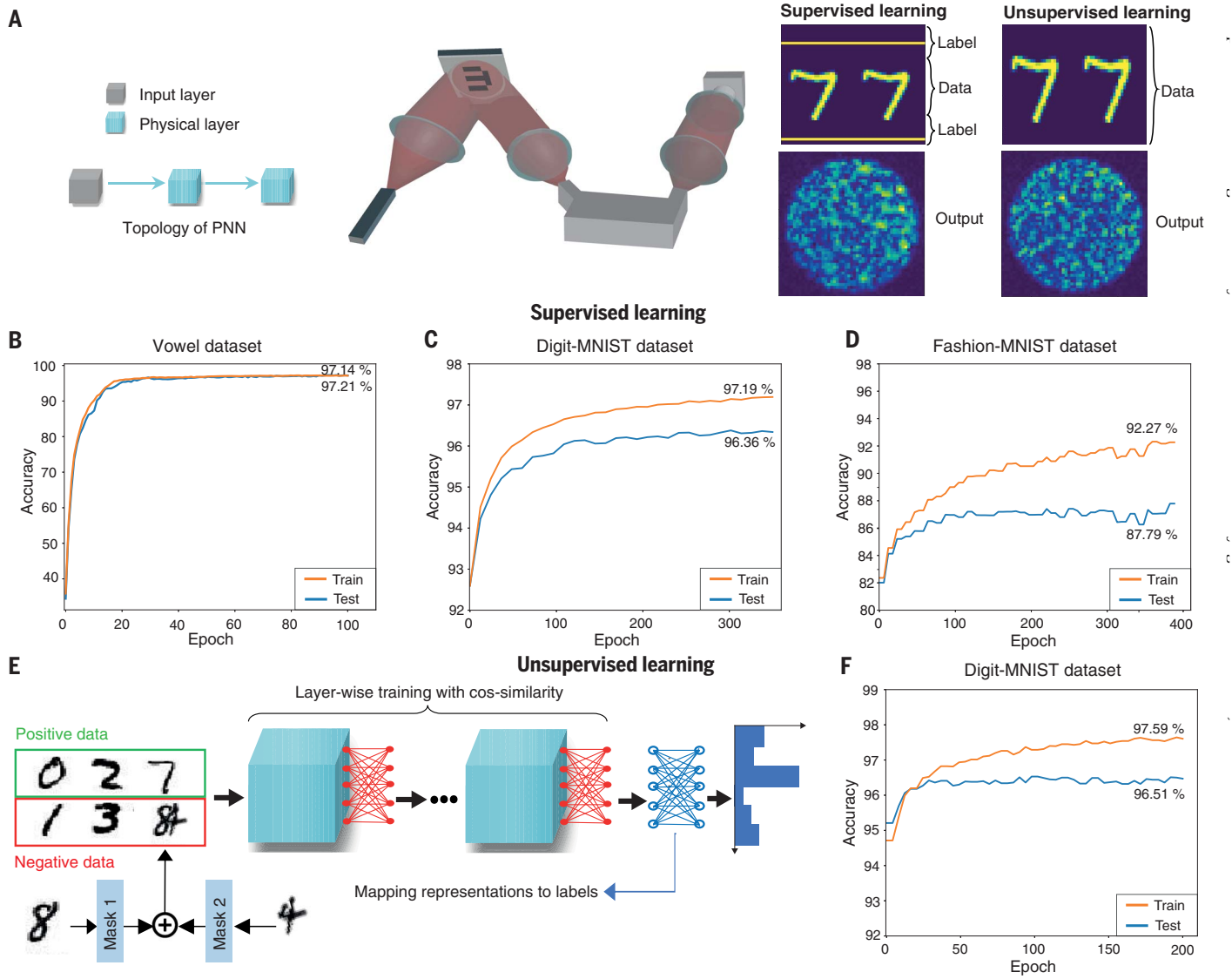


Fig. 4. Optics-PNN. (**A**) The topology of the optics-PNN consists of a two-layer PNN. Each layer comprises an optics-PNN augmented by trainable matrix multiplication. The right panel shows examples of input encoding for supervised and unsupervised contrastive versions along with the corresponding output on a CCD camera for the digit Mnist dataset. (**B** to **D**) The train and test classification accuracy versus training epoch for the vowel (B), digit (C), and fashion Mnist (D) tasks. (**E**) Schematic of the unsupervised version for PNNs (supplementary text, section S2). (**F**) The classification accuracy on the training and test sets versus training epoch for the unsupervised contrastive version of PhyLL on the digit Mnist dataset.

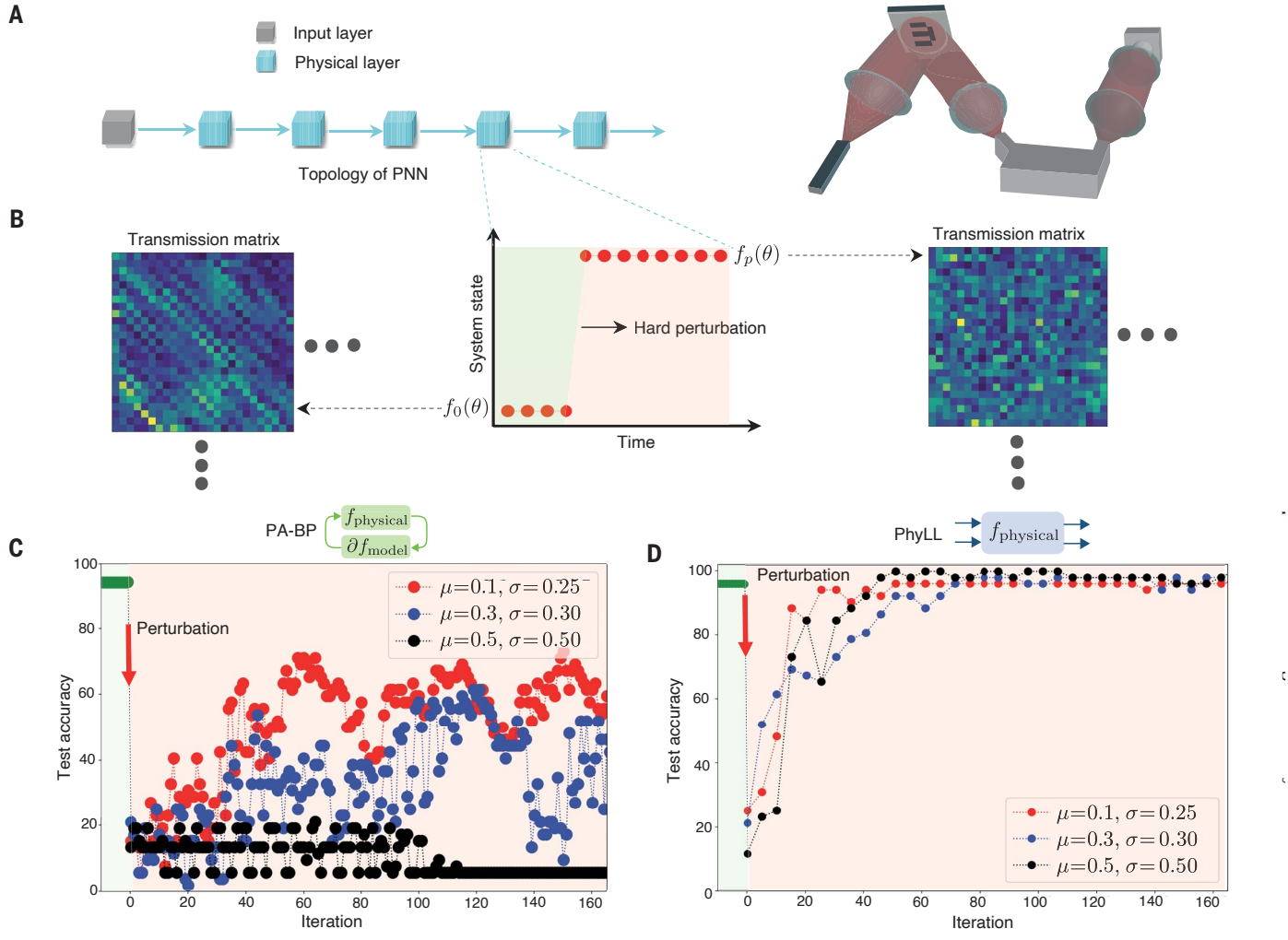


Fig. 5. Robustness of deep PNNs against unpredictable external perturbations. (**A**) A deep PNN consists of six layers of optics-PNN augmented by trainable matrix multiplication. The deep PNN is trained on vowel datasets and is currently in the inference phase. (**B**) Applying hard perturbation by adding Gaussian noise with the mean of μ and standard deviation of σ to the transmission matrix of MMF. (**C** and **D**) A comparison between PA-BP (*22*) (C) and the proposed PhyLL method (D) is presented, with the focus on their ability to recover the classification accuracy after applying perturbation.

and in silico training. A schematic of the aforementioned methods is provided above Fig. 2C. The complete comparison between different hardware-based training methods is provided in the supplementary text, section 2.1.

As demonstrated in Fig. 2C, in silico training performed poorly, reaching only a maximum vowel classification accuracy of ~50%. When there was a gap between the reality and the simulation of a physical system (called the reality-simulation gap), the accuracy of inference would decrease. By contrast, PhyLL succeeded in accurately training the acoustic-PNN, performing similarly to the ideal BP algorithm used as a baseline. The key advantage of PhyLL stems from the execution of both forward passes through the physical hardware rather than simulations.

Microwave massively parametrized chaotic cavity with structural nonlinearity

In the microwave regime, we leveraged a "structural nonlinearity" such that we could implement nonlinear mathematical operations at low power levels with a linear scattering system. Our system consisted of a chaotic cavity that was massively parametrized by covering one of its walls with a programmable metasurface. For each meta-atom and each polarization, the programmable metasurface offered two possible local boundary conditions. Our setup is shown in Fig. 3B and further detailed in the supplementary text, section 2.4. Although the setup resembles that recently used to implement with high fidelity and in situ reprogrammability desired linear transfer functions for signal differentiation (*53*) and routing (*54*), in this case we sought a nonlinear mapping. Hence, we defined the metasurface configuration as the input and the transfer function as the output of our mathematical operation. This relation is in general nonlinear because of the mutual coupling between meta-atoms caused by their proximity and, notably, the reverberation (*55*). We embrace reverberation to maximize the nonlinearity, whereas previous work (*50*) has sought to limit the reverberation to implement a linear transformation with the same input-output definition (see supplementary text, section 2.4, for further discussion).

We randomly grouped our programmable metasurface's 152 degrees of freedom into 40 macropixels because our mathematical operation necessitated 40 input values. We defined our mathematical operation's outputs as the transfer function intensities at 20 decorrelated frequencies within the bandwidth of operation of the programmable metasurface (400 MHz around 5.2 GHz). Thereby, in addition to the structural nonlinearity, we added a readout nonlinearity by considering the transfer function's intensity. To flexibly evaluate the proposed approach, we learned a digital surrogate forward model of the configuration to transfer function intensity mapping (supplementary text, section 2.4). Then, we constructed

the three-layer microwave-PNN shown in Fig. 3A and trained it according to the PhyLL. The training converged after roughly 20 epochs, and the achieved classification accuracy on the unseen test data reached 97.31% (Fig. 3, C to E).

Optical multimode fiber with readout nonlinearity

In the optics example, we used the experimental transfer matrix data of an optical system that comprised an SLM, a scattering medium consisting of a step-index multimode fiber (MMF), and a complementary metal oxide semiconductor (CMOS) camera (Fig. 1D and materials and methods) (*56*). The data were encoded onto the SLM, and after passing through the MMF, the resulting optical intensity was seen on the camera. The physical optical system performed a complex spatial transformation. Although this transformation was linear in the complex domain, the process became nonlinear as a result of the data being encoded onto the phase (SLM) and the subsequent measurement of the intensity squared on the camera.

We used an optics-PNN to perform classification tasks on different datasets: vowel, digit, and fashion Mnist (see Fig. 4; table S1; and supplementary text, section 2.5, for further details). The two-layered optics-PNN achieved high classification accuracy on both the vowel and Mnist dataset. For vowel, we obtained 97.21% and 97.14% accuracy on the training and test sets, respectively. Using only the two-layer optics-PNN, the model achieved 97.19% and 96.36% accuracy on the training and test sets of Mnist, respectively. To implement a deeper optics-PNN, we trained it with six layers on the fashion Mnist dataset, achieving training and test accuracies of 92.27% and 87.79%, respectively (see corresponding results in Fig. 4, B to D). In addition, the unsupervised contrastive learning version of PhyLL is tested on three layers of optics-PNN (Fig. 4E). The model achieved training and test accuracies of 97.59% and 96.51%, respectively (Fig. 4F and table S2). Results were also consistent with a recent preprint for a similar experiment that directly implemented the original forward-forward algorithm, leading to slightly lower efficiency (*57*).

Real-time adaptable learning

In this work, we investigated the robustness of PhyLL in the context of real-time and adaptable learning, where the physical data transformer may undergo changes as a result of the slow dynamics of the physical system during the runtime or external hard perturbations (see also supplementary text, section 2.8). Let us consider a deep optics–PNN with six layers, as depicted in Fig. 5A, which has already been trained on vowel datasets and is currently in the inference phase. The transformation function of each physical system is $f_0(\theta)$, where θ is the physical input. We perturbed the physical systems at a specific time (examples of such perturbations include changes in the MMF state or the positions of lenses or masks, etc.), which results in a change in the transformation function of each physical system from $f_0(\theta)$ to $f_p(\theta)$ (Fig. 5B). To show this, we perturbed the transmission matrix of the optical setup by adding a Gaussian noise with mean μ and standard deviation σ. As observed in Fig. 5D, the test accuracy dropped as expected after applying the perturbation. The question now is whether the training method can restore the accuracy by retraining the optics-PNN. We compared our results with the PA-BP method (*22*), which uses a digital model for the backward pass and the physical system for the forward pass. PA-BP struggled to restore accuracy with increasing perturbation intensity (Fig. 5C). For instance, the test accuracy oscillated around 55% for a small perturbation (red dots in Fig. 5C) and worsened further for more intense perturbations. By contrast, the proposed PhyLL could easily recover accuracy after a few epochs, regardless of the intensity of the perturbation applied (Fig. 5D). This adaptability could be attributed to PhyLL executing both forward passes through the physical hardware rather than digital models. By contrast, the PA-BP method relied on a digital model that lost its accuracy when subjected to hard perturbation, necessitating retraining from scratch or comprehensive hyperparameter tuning.

Discussion

Because of the unprecedented growth in the size of ANNs, such as large language models (LLMs) that are expected to increase unceasingly, the costs of both the training and inference phases of these networks have increased exponentially. Specialized hardware, such as PNNs, have the potential to drastically decrease these costs. Anderson *et al.* (*21*) recently projected an inference–time energy–efficiency advantage of ~8000× compared with that of digital-electronic processors for large-scale future transformer models. The training method proposed in this paper could serve as a viable candidate for training these optical LLMs, potentially offering substantial energy efficiency and speed advantages. We further examine these in the supplementary text, section 2.10. Implementing large-scale LLMs with optics still faces a few challenges, such as the current SLM capacity limited to a few million parameters—far from the billions required. However, there are no fundamental roadblocks to achieving billion-parameter optical architectures and energy-efficient PNNs.

REFERENCES AND NOTES

1. Y. LeCun, Y. Bengio, G. Hinton, *Nature* **521**, 436–444 (2015).
2. L. Deng, D. Yu, *Found. Trends Signal Process.* **7**, 197–387 (2014).
3. M. Asghari, A. V. Krishnamoorthy, *Nat. Photonics* **5**, 268–270 (2011).
4. D. Patterson *et al.*, Carbon Emissions and Large Neural Network Training arXiv:2104.10350 [cs.LG] (2021).
5. G. Wetzstein *et al.*, *Nature* **588**, 39–47 (2020).
6. X. Lin *et al.*, *Science* **361**, 1004–1008 (2018).
7. Z. Wu, M. Zhou, E. Khoram, B. Liu, Z. Yu, *Photon. Res.* **8**, 46–50 (2020).
8. T. W. Hughes, I. A. D. Williamson, M. Minkov, S. Fan, *Sci. Adv.* **5**, eaay6946 (2019).
9. A. Momeni, R. Fleury, *Nat. Commun.* **13**, 2651 (2022).
10. M. Romera *et al.*, *Nature* **563**, 230–234 (2018).
11. J. Grollier *et al.*, *Nat. Electron.* **3**, 360–370 (2020).
12. T. Chen *et al.*, *Nature* **577**, 341–345 (2020).
13. H.-C. Ruiz Euler *et al.*, *Nat. Nanotechnol.* **15**, 992–998 (2020).
14. Y. Shen *et al.*, *Nat. Photonics* **11**, 441–446 (2017).
15. L. Appeltant *et al.*, *Nat. Commun.* **2**, 468 (2011).
16. A. Momeni, X. Guo, H. Lissek, R. Fleury, Physics-inspired Neuroacoustic Computing Based on Tunable Nonlinear Multiple-scattering arXiv:2304.08380 [cs.SD] (2023).
17. A. Momeni, X. Guo, H. Lissek, R. Fleury, in *2022 Sixteenth International Congress on Artificial Materials for Novel Wave Phenomena (Metamaterials 2022)* (IEEE, 2022), pp. 314–316.
18. G. W. Burr *et al.*, *Adv. Phys. X* **2**, 89–124 (2017).
19. D. Marković, A. Mizrahi, D. Querlioz, J. Grollier, *Nat. Rev. Phys.* **2**, 499–510 (2020).
20. M. Prezioso *et al.*, *Nature* **521**, 61–64 (2015).
21. M. G. Anderson, S.-Y. Ma, T. Wang, L. G. Wright, P. L. McMahon, Optical transformers arXiv:2302.10360 [cs.ET] (2023).
22. L. G. Wright *et al.*, *Nature* **601**, 549–555 (2022).
23. M. Nakajima *et al.*, *Nat. Commun.* **13**, 7847 (2022).
24. T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, G. Hinton, *Nat. Rev. Neurosci.* **21**, 335–346 (2020).
25. S. Pai *et al.*, Experimentally realized in situ backpropagation for deep learning in nanophotonic neural networks arXiv:2205.08501 [cs.ET] (2022).
26. X. Guo, T. D. Barrett, Z. M. Wang, A. Lvovsky, *Photon. Res.* **9**, B71–B80 (2021).
27. T. W. Hughes, M. Minkov, Y. Shi, S. Fan, *Optica* **5**, 864–871 (2018).
28. S. Bandyopadhyay *et al.*, Single chip photonic deep neural network with accelerated training arXiv:2208.01623 [cs.ET] (2022).
29. J. F. Miller, S. L. Harding, G. Tufte, *Evol. Intell.* **7**, 49–67 (2014).
30. J. Bueno *et al.*, *Optica* **5**, 756–760 (2018).
31. G. Hinton, The forward-forward algorithm: Some preliminary investigations arXiv:2212.13345 [cs.LG] (2022).
32. H. Zhu, Y. Chen, G. Hu, S. Yu, *Electronics* **12**, 147 (2022).
33. H.-C. Lee, J. Song, SymBa: Symmetric backpropagation-free contrastive learning with forward-forward algorithm for optimizing convergence arXiv:2303.08418 [cs.CV] (2023).
34. R. F. Srinivasan *et al.*, Forward learning with top-down feedback: Empirical and analytical characterization arXiv:2302.05440 [cs.LG] (2023).
35. J. R. Movellan, in *Connectionist Models*, D. S. Touretzky, J. L. Elman, T. J. Sejnowski, G. E. Hinton, Eds. (Elsevier, 1991), pp. 10–17.
36. D.-H. Lee, S. Zhang, A. Fischer, Y. Bengio, in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Proceedings Part I*, A. Appice *et al.*, Eds. (Springer, 2015), pp. 498–515.
37. B. Scellier, Y. Bengio, *Front. Comput. Neurosci.* **11**, 24 (2017).
38. A. N. Gomez *et al.*, *J. Mach. Learn. Res.* **23**, 171 (2022).
39. E. Belilovsky, M. Eickenberg, E. Oyallon, in *Proceedings of the 36th International Conference on Machine Learning*, K. Chaudhuri, R. Salakhutdinov, Eds. (PMLR, 2019), pp. 583–593.
40. Y. Xiong, M. Ren, R. Urtasun, in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin, Eds. (Curran Associates, Inc., 2020), pp. 11142–11153.
41. S. A. Siddiqui, D. Krueger, Y. LeCun, S. Deny, Blockwise self-supervised learning at scale arXiv:2302.01647 [cs.CV] (2023).
42. C. Gallicchio, A. Micheli, L. Pedrelli, *Neurocomputing* **268**, 87–99 (2017).
43. H. Zhou *et al.*, *Light Sci. Appl.* **11**, 30 (2022).
44. M. W. Matthès, P. del Hougne, J. De Rosny, G. Lerosey, S. M. Popoff, *Optica* **6**, 465–472 (2019).
45. A. Silva *et al.*, *Science* **343**, 160–163 (2014).
46. X. Xu *et al.*, *Nature* **589**, 44–51 (2021).
47. A. Babaee, A. Momeni, A. Abdolali, R. Fleury, *Phys. Rev. Appl.* **15**, 044015 (2021).

48. T. Wang *et al*., *Nat. Commun.* **13**, 123 (2022).
49. H. Rajabalipanah, A. Momeni, M. Rahmanzadeh, A. Abdolali, R. Fleury, *Nanophotonics* **11**, 1561–1571 (2022).
50. P. del Hougne, G. Lerosey, *Phys. Rev. X* **8**, 041037 (2018).
51. A. Momeni *et al*., *IEEE Trans. Antenn. Propag.* **69**, 7709–7719 (2021).
52. A. Momeni, M. Safari, A. Abdolali, N. P. Kherani, R. Fleury, *Phys. Rev. Appl.* **15**, 034010 (2021).
53. J. Sol, D. R. Smith, P. Del Hougne, *Nat. Commun.* **13**, 1713 (2022).
54. J. Sol, A. Alhulaymi, A. D. Stone, P. Del Hougne, *Sci. Adv.* **9**, eadf0323 (2023).
55. A. Rabault *et al*., On the tacit linearity assumption in common cascaded models of RIS-parametrized wireless channels arXiv:2302.04993 [cs.IT] (2023).
56. B. Rahmani *et al*., *Nat. Mach. Intell.* **2**, 403–410 (2020).
57. I. Oguz *et al*., *Opt. Lett.* **48**, 5249–5252 (2023).
58. A. Momeni, PhyLL: Backpropagation-free Training of Deep Physical Neural Networks, Github (2023); https://github.com/MomeniAli/PhyLL.
59. A. Momeni, PhyLL: Adding the figures data, Zenodo (2023); https://doi.org/10.5281/zenodo.10075534.

ACKNOWLEDGMENTS

A.M. and M.M. thank X. Guo for useful discussions. B.R. acknowledges that all materials pertaining to the optical experiment presented in this manuscript have been sourced from previously published material that is publicly available and were acquired during his tenure at the Laboratory of Applied Photonics Devices at EPFL. This work has no connection to his current employer in any capacity or form. **Funding:** A.M. and R.F. acknowledge funding from the Swiss National Science Foundation under the Eccellenza grant no. 181232. P.d.H. and R.F. acknowledge funding from the ANR-SNF PRCI program (project "MetaLearn": ANR-22-CE93-0010-01). **Author contributions:** A.M. conceived the idea, designed the computational engine, and carried out both the theoretical and numerical simulations as well as a part of the acoustic experiment. B.R. provided the optics data and interpretation of machine learning results. M.M. carried out the acoustic experiment. P.d.H. carried out the microwave experiment. R.F. supervised the project. All authors contributed to the interpretation of the results and the writing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Materials and methods to evaluate the conclusions in the paper are present in the supplementary materials. All other software and data for running the simulations and experiments are available through Github (*58*). Data underlying the figures are available through Zenodo (*59*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi8474
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S6
References (*60*–*69*)

Submitted 22 May 2023; resubmitted 25 September 2023
Accepted 7 November 2023
Published online 23 November 2023
10.1126/science.adi8474

ANCIENT DNA

The history of Coast Salish "woolly dogs" revealed by ancient genomics and Indigenous Knowledge

Audrey T. Lin[1,2]*, Liz Hammond-Kaarremaa[1,3]*, Hsiao-Lei Liu[1], Chris Stantis[1,4], Iain McKechnie[5], Michael Pavel[6], Susan sa'hLa mitSa Pavel[6,7,8], Senaqwila Seǹákw Wyss[9], Debra qwasen Sparrow[10], Karen Carr[11], Sabhrina Gita Aninta[12], Angela Perri[13,14], Jonathan Hartt[15], Anders Bergström[16,17], Alberto Carmagnini[18], Sophy Charlton[19,20], Love Dalén[21,22,23], Tatiana R. Feuerborn[24,25], Christine A. M. France[26], Shyam Gopalakrishnan[24], Vaughan Grimes[27], Alex Harris[25], Gwénaëlle Kavich[26], Benjamin N. Sacks[28,29], Mikkel-Holger S. Sinding[30], Pontus Skoglund[16], David W. G. Stanton[18,31], Elaine A. Ostrander[25], Greger Larson[19], Chelsey G. Armstrong[15], Laurent A. F. Frantz[12,18], Melissa T. R. Hawkins[32], Logan Kistler[1]*

Ancestral Coast Salish societies in the Pacific Northwest kept long-haired "woolly dogs" that were bred and cared for over millennia. However, the dog wool–weaving tradition declined during the 19th century, and the population was lost. In this study, we analyzed genomic and isotopic data from a preserved woolly dog pelt from "Mutton," collected in 1859. Mutton is the only known example of an Indigenous North American dog with dominant precolonial ancestry postdating the onset of settler colonialism. We identified candidate genetic variants potentially linked with their distinct woolly phenotype. We integrated these data with interviews from Coast Salish Elders, Knowledge Keepers, and weavers about shared traditional knowledge and memories surrounding woolly dogs, their importance within Coast Salish societies, and how colonial policies led directly to their disappearance.

Dogs were introduced to the Americas from Eurasia via northwestern North America ~15,000 years ago and have been ubiquitous in Indigenous societies of the Pacific Northwest (PNW) for millennia (*1*–*4*). Coast Salish peoples in the Salish Sea region (Fig. 1A) kept multiple different types of dogs: hunting dogs, village dogs, and "woolly dogs" with a thick woolen undercoat that was shorn for weaving (*4*, *5*). Dog-wool blankets, often blended with mountain goat wool, waterfowl down, and plant fibers such as fireweed and cattail fluff, were prestigious cultural belongings (*6*–*8*). Woolly dogs, known as "sqwemá:y," "ske'-ha," "sqʷəméy," "sqʷbaý," and "q'əbəł" in some Coast Salish languages (*9*), were emblems of some communities, as depicted in a 19th-century Skokomish/Twana basket (Fig. 1B) (*10*).

The first comprehensive book on Salish weaving (*11*) scrutinized most Coast Salish woven blankets in museums around the world, questioning if any contained primarily dog wool and disputing the fiber's spinnability. More-recent proteomic analysis of 19th-century blankets confirmed the use of dog wool in Coast Salish weaving (*12*). In addition, zooarchaeological remains thought to be from woolly dogs have been found in dozens of archaeological sites in Coast Salish territories beginning ~5,000 years before present (B.P.) (*2*, *4*) (Fig. 1A). The last Coast Salish woolly dogs likely lived in the late 19th and early 20th centuries (*5*, *13*). Later photographs and records referring to woolly dogs extend into the 20th century, but these examples likely reflect mixed ancestry or non-Indigenous breeds (*9*).

The decline in dog-wool weaving has previously been attributed to the proliferation of machine-made blankets by British and American trading companies in the early 19th century (*11*, *13*). However, this explanation ignores the cultural importance of woolly dogs, as reflected through their enduring use by weavers, particularly for high-status items such as regalia (*7*, *14*). Given their role in Coast Salish societies, it is unlikely that the entire dog-wool tradition would have been abandoned simply because of the ready availability of imported textiles. Furthermore, this explanation ignores weavers' efforts to maintain culturally relevant practices in the face of settler colonialism. The use of blankets and robes served not only a functional purpose but also a spiritually protective role in Coast Salish cultures. Wearing a ceremonial blanket was spiritually transformative because it intertwined the creator of the blanket, the wearer, and the community (*13*–*15*).

The only known pelt of an extinct Coast Salish woolly dog is of "Mutton," a dog cared for by naturalist and ethnographer George Gibbs during the Northwest Boundary Survey (1857–1862). According to Gibbs's field journal and Smithsonian ledgers [National Museum of Natural History (USNM) A4401 to A4425], Mutton became ill and died in late 1859 (*9*, *15*). His pelt and lower leg bones are housed at the Smithsonian Institution (USNM 4762) (figs. S2 and S4).

In this study, we combined genomic analysis, ethnographic research, stable isotope and zooarchaeological analysis, and archival records to investigate this iconic dog's history, including ancestry, the genetic underpinnings of woolliness, and their ultimate decline. We sequenced Mutton's nuclear genome to a mean 3.4× depth of coverage and, for comparison, a nonwoolly village dog (figs. S3 and S5) from the nearby Semiahmoo Bay region to low erage (0.05×; "SB dog" hereafter, USNM 3[illegible], collected 1858). For additional genomic context, we increased the coverage of an ancient dog from Port au Choix, Newfoundland [AL3194; 4020 calibrated years B.P. (cal B.P.)] (*3*), from 1.9× to 11.9×, and sequenced the genome of an ancient dog from Teshekpuk Lake, Alaska (ALAS_015; 3763 BP; 1.23×); three modern coyotes; and 59 modern dogs representing 21 breeds (data S1). We also undertook $\delta^{13}C$ and $\delta^{15}N$ stable-isotope analysis of Mutton and the SB dog to test for substantial differences in their dietary life histories. Lastly, we interviewed seven Coast Salish Elders, Knowledge Keepers, and wool weavers about family histories and traditional knowledge surrounding woolly dogs to provide a cultural framework for interpreting the genomic analyses (*9*). The interviewees span several Coast Salish communities, including Stó:lō, Squamish, Snuneymuxw, and Musqueam Nations in British Columbia (BC) and Suquamish and Skokomish/Twana in Washington.

Woolly dog origins

Throughout northwestern North America there are numerous oral histories and origin stories

[1]Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. [2]Richard Gilder Graduate School, American Museum of Natural History, New York, NY, USA. [3]Vancouver Island University, Nanaimo, BC, Canada. [4]Department of Geology and Geophysics, University of Utah, Salt Lake City, UT, USA. [5]Department of Anthropology, University of Victoria, Victoria, BC, Canada. [6]Twaṇa/Skokomish Indian Tribe, Skokomish Nation, WA, USA. [7]Coast Salish Wool Weaving Center, Skokomish Nation, WA, USA. [8]The Evergreen State College, Olympia, WA, USA. [9]Skwxwú7mesh Úxwumixw (Squamish Nation), North Vancouver, BC, Canada. [10]Musqueam First Nation, Vancouver, BC, Canada. [11]Karen Carr Studio, Silver City, NM, USA. [12]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. [13]Department of Anthropology, Texas A&M University, College Station, TX, USA. [14]Chronicle Heritage, Phoenix, AZ, USA. [15]Department of Indigenous Studies, Simon Fraser University, Burnaby, BC, Canada. [16]Ancient Genomics Laboratory, The Francis Crick Institute, London, UK. [17]School of Biological Sciences, University of East Anglia, Norwich, UK. [18]Palaeogenomics Group, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany. [19]PalaeoBARN, School of Archaeology, University of Oxford, Oxford, UK. [20]BioArCh, Department of Archaeology, University of York, York, UK. [21]Centre for Palaeogenetics, Stockholm, Sweden. [22]Department of Zoology, Stockholm University, Stockholm, Sweden. [23]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden. [24]Center for Evolutionary Hologenomics, The Globe Institute, University of Copenhagen, Copenhagen, Denmark. [25]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [26]Museum Conservation Institute, Smithsonian Institution, Suitland, MD, USA. [27]Department of Archaeology, Memorial University of Newfoundland, St. Johns, NL, Canada. [28]Mammalian Ecology and Conservation Unit, Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA. [29]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA. [30]Department of Biology, University of Copenhagen, Copenhagen, Denmark. [31]Cardiff School of Biosciences, Cardiff University, Cardiff, UK. [32]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA.
*Corresponding author. Email: linat@si.edu (A.T.L.); liz.hammond-kaarremaa@viu.ca (L.H.-K.); kistlerl@si.edu (L.K.)

involving the woolly dog. Skokomish/Twana Elder Michael Pavel reports that in a former time, when all beings including woolly dogs were recognized as relatives, all were "people" and were as family. High-status Qw'ó:ntl'an women are an example of those who trace their lineages from the woolly dog at a time when all beings were one family (*16*). According to Pavel: "And out of [the origin story], [woolly dogs] were given the gift of the wool, and they were able to teach the women how to gather the wool, how to process the wool, how to spin the wool, and how to weave with the wool" (*9*).

Early colonial explorers and scholars speculated that woolly dogs originated in Japan (*17*) or were recently introduced to the Coast Salish by the Dene people from their homelands in northern boreal Canada (*18*). However, zooarchaeological remains of morphologically distinct dogs in Coast Salish territories suggest that woolly dog husbandry was present for ~5000 years before European colonization (*2*, *4*). Furthermore, longstanding oral histories and traditional knowledge hold that woolly dogs have been part of Coast Salish society for millennia (*9*).

To test whether Mutton has precolonial or settler dog ancestry, we first compared his mitochondrial genome with 207 ancient and modern dogs from a global sampling. Mutton carries the A2b mitochondrial DNA (mtDNA) haplotype, which emerged after dogs initially arrived from Eurasia (*3*). Most of this mtDNA lineage of so-called precolonial dogs (PCDs) disappeared after European colonization (*3*, *19*, *20*). Mutton's nearest mtDNA neighbor is an ancient dog (PRD10, ~1,500 B.P.) from Prince Rupert Harbour, BC (Fig. 2A and fig. S16). PRD10 is the only archaeological dog from the PNW in the mtDNA dataset, and this similarity reflects the deep roots of Mutton's maternal ancestry in the region. A pair of modern and ancient (~620 B.P.) dogs from Alaska form a sister clade of the Mutton-PRD10 grouping, further underscoring the long-term maternal population structure in northwestern North America. By contrast, the SB dog carries an A1a haplotype, which is similar to that of most modern European dogs and is the most common present-day haplotype worldwide (found in 64 out of 207 dogs in our analysis) (*21*).

To place a timeframe on the divergence of Mutton's maternal lineage, we performed a molecular-clock analysis on the mitochondrial phylogeny (data S1). The results suggest a mitochondrial common ancestor estimated between 4776 and 1853 years B.P. for the subclade containing Mutton, PRD10, and the two Alaskan dogs (95% highest posterior density; Fig. 2A and fig. S16). Although we are limited by the analysis of a single individual, this timing is generally consistent with the increasing occurrence of small-sized woolly dog zooarchaeological remains in the regions surrounding the Salish Sea (*2*).

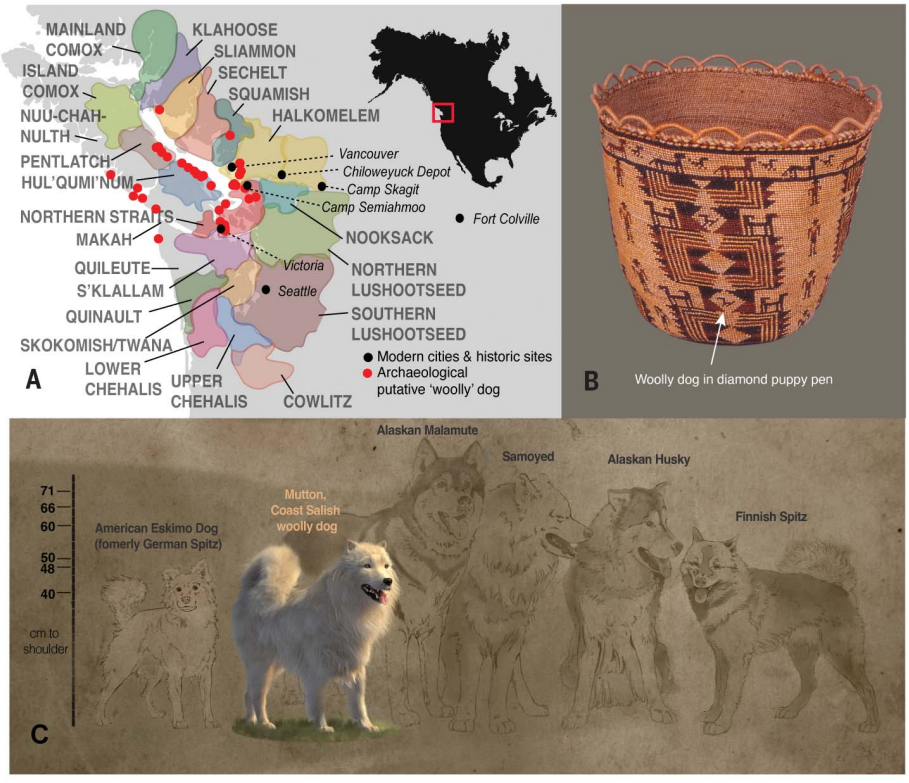


Fig. 1. Domestic dogs in the culture and society of Indigenous Coast Salish peoples. (**A**) Coast Salish ancestral lands include the inner coastal waterways of the Salish Sea in southwest British Columbia and Washington State. Archaeological woolly dog data are from (*2*). Distribution of the Coast Salish languages in the 19th century are as indicated by colored areas. [The map is modified from https://commons.wikimedia.org/wiki/File:Coast_Salish_language_map.svg and licensed under CC BY-SA 4.0.] (**B**) Woven Skokomish/Twana basket with woolly dog iconography, depicted with upturned tails. Woolly dog puppies are inside pens represented by diamond shapes (*10*) [courtesy of Burke Museum, catalog no. 1-507]. (**C**) Forensic reconstruction of a woolly dog based on Mutton's pelt measurements and archaeological remains (*9*). Sketches of Arctic and spitz dog breeds are shown for scale and comparison of appearance and do not imply a genetic relationship.

To assess Mutton's nuclear ancestry, we analyzed 217 globally distributed ancient and modern dogs. Outgroup-f_3 statistics reveal that Mutton carries substantially greater shared genetic drift with PCDs than with any other dogs, specifically, with the archaeological remains of a dog from Port au Choix, Newfoundland (4020 cal B.P.), and from Weyanoke Old Town, Virginia (~1,000 B.P.) (Fig. 2B and fig. S17). Because Mutton lived after European colonization and waves of precolonial dog introductions (*3*, *21*), we tested for gene flow from introduced lineages using D-statistics. We found that European breeds yielded strongly positive D-statistics, indicating that Mutton's non-PCD ancestry most likely stemmed from introduced European dogs (Fig. 2C).

To refine these results, we used f_4-ratio tests with six modern European breeds (Chinese Crested dog, English Cocker Spaniel, Dalmatian, German Shepherd, Lagotto Romagnolo, and Portuguese Water Dog), estimating that Mutton had 84% PCD ancestry and 16% European ancestry (11.9 to 19.9% 2 SE range; Fig. 2D). The f_4-ratio test may slightly overestimate Mutton's European ancestry if the true contributor of this ancestry was equally related (an outgroup) to the two European breeds in the tests. However, estimates across all permutations are broadly consistent (Fig. 2D and fig. S18), suggesting European ancestry roughly on the order of one great-grandparent in Mutton's background. By contrast, outgroup-f_3 statistics indicate that the contemporaneous SB dog appears highly admixed, showing the greatest similarity to ancient dogs from Siberia and Alaska (fig. S17). The distribution of PCD versus European ancestry tracts in Mutton can provide some additional insight into the timing of admixture. Although this method is imprecise because of recent admixture and the scarcity of PCD source–population data, we estimate that Mutton's European admixture occurred 10.8 ± 4.9 generations before (1 SE). Assuming a 3-year generation time, this analysis suggests admixture ~32 years before Mutton's birth, consistent with postcolonial admixture (*9*).

To test for dietary differences between Mutton and the SB dog, we performed stable isotope analysis of $\delta^{13}C$ and $\delta^{15}N$ on bone collagen and

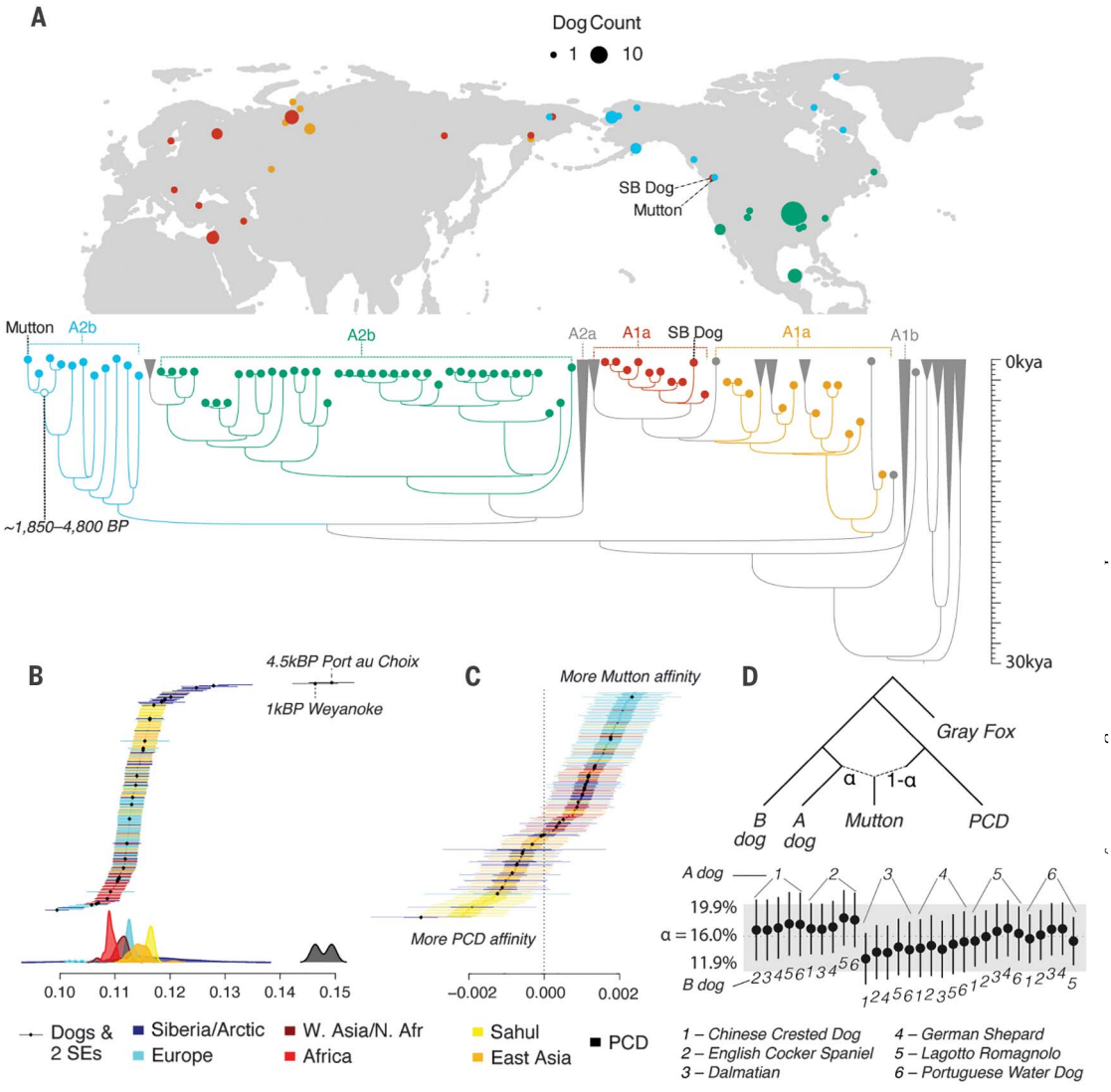


Fig. 2. Genetic ancestry of woolly dogs. (**A**) mtDNA tree of 207 dogs with A2b (Mutton) and A1a (SB Dog) haplotypes expanded. The map points correspond to colored tree tips for the most similar archaeological and historic dog mtDNAs, highlighting the subclades of interest and the broader haplotypes. Samples used are listed in data S1. (**B**) Outgroup-f_3 statistics (f_3(Gray Fox; Mutton, B)) or estimation of shared drift between Mutton and 229 other dogs revealed that Mutton has the highest similarity to PCDs. Black-point estimates indicate ancient genomes. (**C**) D-statistics (((PCD, Mutton), Test Dog), Gray Fox) consistent with gene flow into Mutton's background, with European breeds appearing the most likely contributors to Mutton's non-PCD ancestry. (**D**) f_4-ratio tests (f_4(A, Out; Mutton, AL3194-Port au Choix): f_4(A, Out; B, AL3194-Port au Choix)) to estimate the proportion of European settler-dog ancestry in Mutton's background, performed by using six modern European breeds as proxies for Mutton's European ancestry component.

hair keratin. The SB dog has high $\delta^{13}C$ and $\delta^{15}N$ values similar to those of archaeological dogs from the PNW (*22*), indicating a traditional marine-based diet (figs. S13 and S14). Mutton's isotope values reveal a more terrestrial and complement C_3 component–rich diet, likely reflecting Mutton's life and travels with Gibbs from an early age (fig. S14, B and C, and fig. S15) (*9*).

The persistence of a high proportion of post-colonial PCD ancestry may reflect concerted efforts by Coast Salish peoples to maintain the breed against the pressure of gene flow from nonnative dogs. Mutton lived near the end of traditional woolly dog husbandry (*5*, *9*, *13*). Although he had mixed ancestry, Mutton's background is dominated by PCD ancestors, as compared with that of the contemporaneous SB dog. This finding may indicate careful reproductive management to maintain woolly dogs' distinct genetic makeup and phenotype until their decline. Mutton's fraction of European ancestry also highlights the turbulent cultural moment at the time Mutton lived and illustrates how interbreeding with settler-introduced dogs could have threatened the survival of woolly dogs.

The influence of people on the woolly dog genome

Woolly dogs were treated as beloved extended family members. According to Debra qwasen Sparrow, a Musqueam Master Weaver, her grandfather [Ed Sparrow (1898–1998)] told her that "every village had [woolly dogs], that they were like gold because they were mixed with the mountain goat and then rove and spun" (*9*). Dogs also comprised a form of wealth and status for Coast Salish women, who carefully managed the dogs to maintain their woolly coats, isolating them on islands or in pens to strictly manage their breeding (*9*, *17*, *23*). Island names often reflect their connection with dogs, such as "sqwiqwmi" ("Little Dog") village on Cameron Island in Nanaimo, Snuneymuxw territory, BC. The prevention of interbreeding wool dogs with hunting or village dogs was critical for maintaining their distinct hair characteristics: soft guard hairs with an unusually long, crimpy undercoat (fig. S2), which was highly spinnable and could be made into warm blanket yarn. These management practices likely contributed to Mutton's PCD ancestry long after the onset of settler colonialism.

Long-term husbandry for woolly hair likely limited woolly dogs' effective population size, which would be reflected in nucleotide diversity and thus in Mutton's heterozygosity. We found that Mutton's heterozygosity is in the lowest range of living breeds (n = 51) and village dogs (n = 42) downsampled to the same coverage (Fig. 3A). Additionally, runs of homozygosity (ROH) better reflect recent demography than global heterozygosity. Using an ROH method optimized for low coverage (*9*, *24*), we estimate that 15.7% of Mutton's genome is in ROH of 2.5 mega–base pairs (Mbp) or greater, again in the range of modern breeds. The ancient Port au Choix dog also has low genomic heterozygosity and 11.3% ROH, so Mutton's low

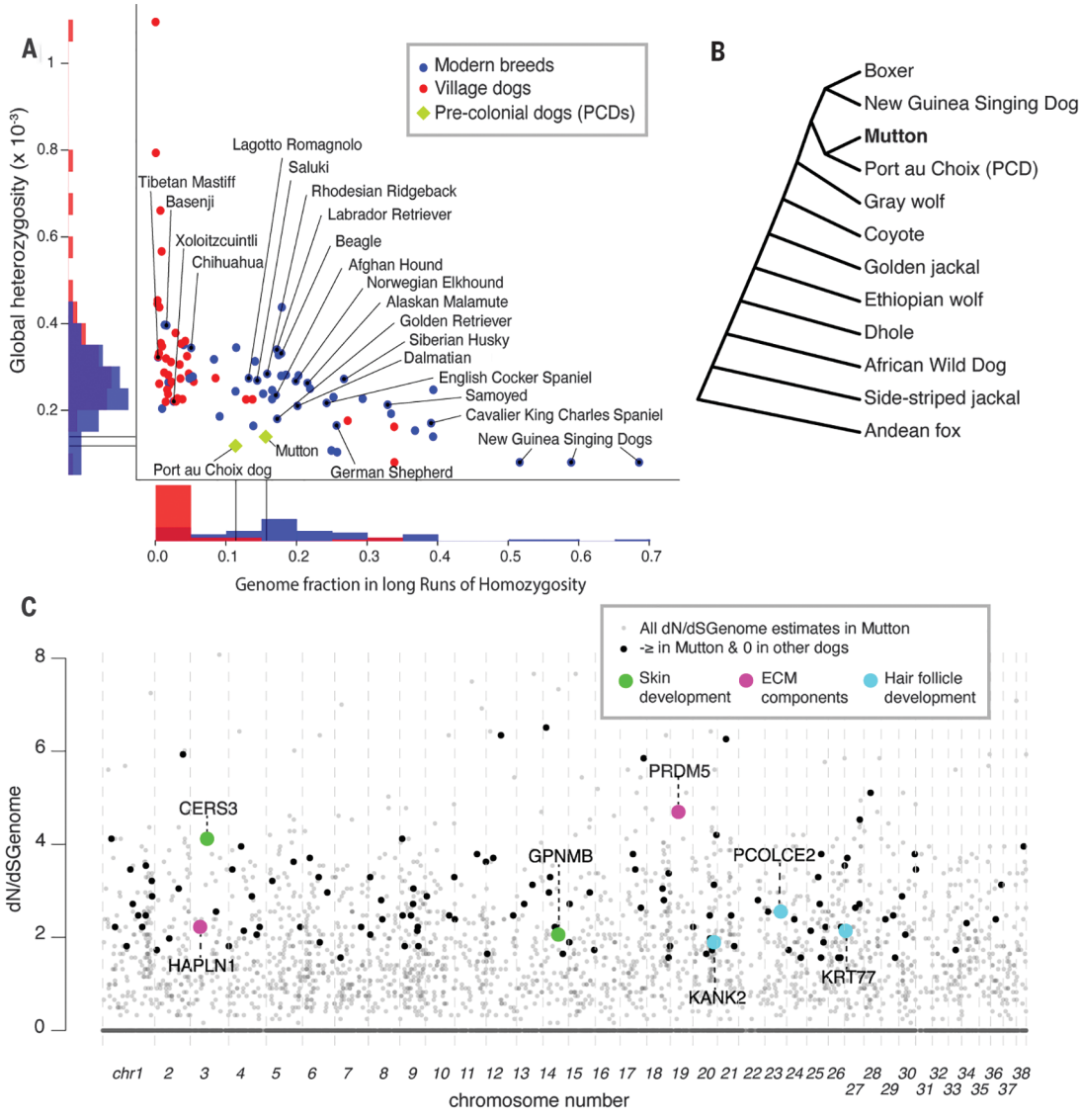


Fig. 3. Genomic outcomes of management and selection. (**A**) Global heterozygosity and long runs of homozygosity over transversions in Mutton compared with modern dogs and the ancient Port au Choix dog. All dogs have been downsampled to Mutton's coverage level for analysis. (**B**) Tree schematic used in dN/dS analysis to identify genes under selection in Mutton compared with other canids. The branching order is based on (*50*). dN/dS_{genome} estimates were done separately including one of the four dogs plus all other canids. Genes with elevated dN/dS_{genome} values in multiple dogs could reflect more ancient shared selection before the separation of the woolly dog lineage. Therefore, likely candidates for selection in woolly dogs were conservatively assessed where $dN/dS_{genome} > 1.5$ in Mutton (*9*) but dN = 0 in the other three dogs, including one PCD. (**C**) Genes with an excess of nonsynonymous mutations in Mutton. Black points are the 125 selection candidates identified on the basis of $dN/dS_{genome} \geq 1.5$ in Mutton but dN = 0 in three other dogs, including one PCD (*9*). Several genes with high dN/dS_{genome} in Mutton (shown in gray) are excluded as selection candidates because they carry at least one nonsynonymous mutation in other dogs. This approach is designed to conservatively highlight genes in which selection is more likely specific to Mutton's lineage rather than during dog domestication or in the common ancestors of PCDs. Candidate genes discussed in text are indicated.

heterozygosity may partly reflect shared demographic history from a small PCD founding population (Fig. 3A). Because of recent European admixture, Mutton's genome is inevitably more heterozygous than that of his recent woolly dog ancestors.

To search for evidence of genetic mechanisms for woolliness, we used maximum likelihood–based estimation of the enrichment of nonsynonymous mutations (dN/dS) (ratio of nonsynonymous to synonymous mutations) observed within Mutton's coding regions (*9*). We evaluated 11,112 genes with sufficient sequence coverage for all dogs and outgroups (data S1) and restricted selection-candidate identification to genes with elevated dN/dS in Mutton but lacking any nonsynonymous mutations in three other dogs, including one PCD (Fig. 3B). Although power to detect selection is fundamentally limited with only a single genome, we identified a candidate set of genes with high lineage-specific dN/dS values. We identified 125 genes as candidates for positive selection in woolly dogs (data S2). Among these, 28 have plausible links to hair growth and follicle regeneration according to a model of the hair-growth cycle (fig. S12) and are associated with cell replication, proliferation, the formation of extracellular matrix components, vascularization, and related processes (*25*–*31*) (Fig. 3C and data S3).

Candidate selection genes in Mutton include *KANK2*, a steroid-signaling regulator responsible for hereditary diseases of the hair shaft in humans (*32*). A distinct nonsynonymous mutation in Mutton lies in the amino acid adjacent to the *KANK2* mutation, causing a "woolly" hair phenotype in humans (*32*). *KRT77* is a member of the keratin gene family responsible for the structural integrity of cells in the epithelium and hair follicles. Mutations in keratin genes are linked to curly-hair phenotype in other dogs, rats, and mice (*31*), and to woolly hair and hereditary hair loss in humans (*26*, *30*); and multiple *KRT* genes underwent selection in woolly mammoths (*25*). *CERS3*, *PRDM5*, and *HAPLN1* are associated with maintaining the integrity of the skin or connective tissue in humans (*27*, *28*). *GPNMB* is involved in multiple cellular functions in the epidermis, potentially mediating pigmentation (*29*). We also manually evaluated 15 specific variants from previous literature that are linked with hair characteristics in living dog breeds (data S4). Apart from a widespread *FGF5* mutation conferring long hair (*33*, *34*), Mutton showed the ancestral allele in all cases with data present (data S4),

illustrating the independent origins of woolly dogs' distinct phenotype.

The impact of colonialism on the iconic breed's disappearance

Woolly dogs' decline throughout the 19th century is not fully understood. The narrative that the influx of trade blankets into the region led to the abandonment of woolly dog husbandry oversimplifies a complex scenario. By 1857 (a year before Mutton's birth) in Stó:lō territory, where Mutton was most likely acquired, the settler population consisted of only a few dozen permanent settlers at Fort Langley (*35*, *36*). The following year, more than 33,000 miners arrived at present-day BC during the 1858 Fraser River Gold Rush. This large-scale migration set off conflicts between miners, colonial governments, and Indigenous peoples. Indigenous populations declined by an estimated two-thirds between 1830 and 1882 (*37*). Smallpox epidemics—almost one every generation from the 1700s to 1862 (*38*)—are estimated to have killed more than 90% of Indigenous people in some villages across BC (*38*), along with steady depopulation due to other introduced diseases such as mumps, tuberculosis, and influenza (*37*).

Survival of woolly dogs depended upon the survival of their caretakers. In addition to disease, expanding colonialism increased cultural upheaval, displacement of Indigenous peoples, and a diminished capacity to manage the breed. Policies targeted Indigenous governance and inherent rights, resulting in the deliberate disenfranchisement and criminalization of Indigenous cultural practices (*39*). Indigenous women, the caretakers of woolly dogs and weaving knowledge, were specifically targeted. Missionization efforts reduced women's roles in society, and legislation such as the Indian Act (1876) explicitly prohibited women from participating in local governance, denied women basic property rights, and restricted their movement (*39*). In the 20th century, transference of cultural knowledge was further disrupted by mandatory residential schooling designed to remove children from their families and suppress culture (*40*).

Through these compounding waves of colonialism, the transmission of important knowledge relating to woolly dog husbandry and hair processing, spinning, and weaving was interrupted. Stó:lō Elder Rena Point Bolton, 95 years old in 2022, recalls how Th'etsimiya, her great-grandmother, had kept woolly dogs, but was forced to give them up: "They were told they couldn't do their cultural things. There was the police, the Indian Agent and the priests. The dogs were not allowed. She had to get rid of the dogs" (*9*). The dogs represented high status and traditional practices that threatened British and later Canadian dominion and as such were removed through policies of assimilation (*40*–*42*). The weaving traditions were not completely lost, because many cultural teachings and types of expertise were carried on in secret. Bolton said: "Our people were not allowed to spin on shxwqáqelets [traditional spindle whorls]. They could spin on a European one but not on the shxwqáqelets. They couldn't use their looms, and they would take them out and burn them or they would give them to museums or collectors … The generation that was there when the Europeans came and colonized us, that's where it ended, and there [were] just a few people who went underground. And my grandmother and my mother were two of them" (*9*).

A growing body of research demonstrates how peoples of the PNW cared for and managed their ancestral lands, cultivating diverse and highly localized plants and marine foods (*43*–*45*). Woolly dogs may have also been similarly localized and diverse. We focused on Coast Salish dogs, but non-Salish peoples in the PNW also kept woolly dogs. For example, Nuu-chah-nulth peoples of western Vancouver Island kept a different wool dog that was reportedly bigger and had coats of different colors, including brown, spotted, black, gray, or white (*46*–*48*). These differences could be population-specific, or they could be a result of widespread phenotypic diversity, as noted by explorers in the 18th and 19th centuries (*17*), reflecting trade among the different Indigenous communities.

Weaving and woolly dogs are intertwined in Coast Salish culture and society, which cannot be separated from the long-time management of their ancestral homelands. Weavers, artists, and Elders continue to promote the renewal of traditional or customary weaving knowledge and practices. Artist Eliot Kwulasultun White-Hill (Snuneymuxw) said (*9*): "It starts to unravel, in a way, people's understanding of us as a hunter gatherer society … Our relationship with the woolly dogs, our relationship with the camas patches and the clam beds, the way that we tended the land and tended the forests … these all show the systems in place that are far more complex than what people take for granted about Coast Salish culture."

REFERENCES AND NOTES

1. D. Fedje, Q. Mackie, D. McLaren, B. Wigen, J. Southon, *Quat. Sci. Rev.* **272**, 107221 (2021).
2. I. McKechnie, M. L. Moss, S. J. Crockford, *J. Anthropol. Archaeol.* **60**, 101209 (2020).
3. M. Ní Leathlobhair *et al.*, *Science* **361**, 81–85 (2018).
4. S. J. Crockford, *Osteometry of Makah and Coast Salish Dogs* (Archaeology Press, Simon Fraser University, 1997).
5. R. Schulting, *Can. J. Archaeol.* **18**, 57–76 (1994).
6. W. H. Dall, G. Gibbs, J. W. Powell, *Tribes of the Extreme Northwest, and Tribes of Western Washington and Northwestern Oregon*, vol. I of *Contributions to North American Ethnology* series (Cosimo Classics, 1877).
7. W. Suttles, in *Indian Art Traditions of the Northwest Coast*, R. L. Carlson, Ed. (Archaeology Press, Simon Fraser University, 1982), p. 70.
8. H. G. Barnett, *The Coast Salish of British Columbia* (University of Oregon, 1955), vol. 4 of *University of OregonMonographs: Studies in Anthropology*.
9. Materials and methods are available as supplementary materials.
10. "Burke Museum Record," Burke Museum basketry exhibition, https://www.burkemuseum.org/static/baskets/idgame/dreport.html.
11. P. Gustafson, *Salish Weaving* (Douglas & McIntyre, 1980).
12. C. Solazzo *et al.*, *Antiquity* **85**, 1418–1432 (2011).
13. R. L. Barsh, J. M. Jones, W. Suttles, in *Proceedings of the 9th Conference of the International Council of Archaeozoology, Durham, August 2002*, L. M. Snyder, E. A. Moore, Eds. (Oxbow Books, 2006), pp. 2–11.
14. L. H. Tepper, J. George, W. Joseph, *Salish Blankets* (Univ. of Nebraska Press, 2017).
15. G. Gibbs, *Journal, Northwest Boundary Survey, 1857–1862*, 1859; https://doi.org/10.5962/bhl.title.97030.
16. K. T. Carlson, in *A Stó:lō-Coast Salish Historical Atlas*, K. Carlson, A. J. McHalsie, Eds. (Douglas & McIntyre, 2001), p. 25.
17. J. K. Lord, *The Naturalist in Vancouver Island and British Columbia* (R. Bentley, 1866).
18. F. W. Howay, *The Washington Historical Quarterly* **9**, 83–92 (1918).
19. A. Bergström *et al.*, *Science* **370**, 557–564 (2020).
20. S. Castroviejo-Fisher, P. Skoglund, R. Valadez, C. Vilà, J. A. Leonard, *BMC Evol. Biol.* **11**, 73 (2011).
21. C. Ameen *et al.*, *Proc. Biol. Sci.* **286**, 20191929 (2019).
22. D. Hillis, I. McKechnie, E. Guiry, D. E. St Claire, C. T. Darimont, *Sci. Rep.* **10**, 15630 (2020).
23. M. Eells, *The Indians of Puget Sound: The Notebooks of Myron Eells*, G. P. Castile, Ed. (Whitman College, 1985).
24. K. G. Daly *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2100901118 (2021).
25. D. Díez-del-Molino *et al.*, *Curr. Biol.* **33**, 1753–1764.e4 (2023).
26. Y. Shimomura, M. Wajid, L. Petukhova, M. Kurban, A. M. Christiano, *Am. J. Hum. Genet.* **86**, 632–638 (2010).
27. F. P. W. Radner *et al.*, *PLOS Genet.* **9**, e1003536 (2013).
28. E. M. M. Burkitt Wright *et al.*, *Am. J. Hum. Genet.* **88**, 767–777 (2011).
29. K. B. Biswas *et al.*, *Sci. Rep.* **10**, 4930 (2020).
30. N. Wasif *et al.*, *Hum. Genet.* **129**, 419–424 (2011).
31. S. Harel, A. M. Christiano, *J. Invest. Dermatol.* **132**, 2315–2317 (2012).
32. Y. Ramot *et al.*, *J. Med. Genet.* **51**, 388–394 (2014).
33. C. Dierks, S. Mömke, U. Philipp, O. Distl, *Anim. Genet.* **44**, 425–431 (2013).
34. E. Cadieu *et al.*, *Science* **326**, 150–153 (2009).
35. J. R. Gibson, *Farming the Frontier: The Agricultural Opening of the Oregon Country, 1786–1846* (Univ. of British Columbia Press, 1985).
36. K. Carlson, in *A Stó:lō-Coast Salish Historical Atlas*, K. Carlson, A. J. McHalsie, Eds. (Douglas & McIntyre, 2001), pp. 76–83.
37. K. T. Carlson, in *A Stó:lō-Coast Salish Historical Atlas*, K. Carlson, A. J. McHalsie, Eds. (Douglas & McIntyre, 2001), pp. 92–93.
38. R. Boyd, *BC Stud.* **101**, 5–40 (1994).
39. B. Lawrence, *Indians and Others: Mixed-Blood Urban Native Peoples and Indigenous Nationhood* (Univ. of Nebraska Press, 2004).
40. E. Hanson, D. P. Gamez, A. Manuel, The Residential School System (Indigenous Foundations, 2020); https://indigenousfoundations.arts.ubc.ca/the_residential_school_system/.
41. R. Fisher, *Contact and Conflict: Indian-European Relations in British Columbia, 1774–1890* (UBC Press, ed. 2, 1992).
42. J. S. Lutz, *Makúk: A New History of Aboriginal-White Relations* (UBC Press, Vancouver, 2008).
43. C. G. Armstrong, J. Earnshaw, A. C. McAlvay, *J. Archaeol. Sci.* **143**, 105611 (2022).
44. D. Lepofsky *et al.*, *Ecosystems* **24**, 248–260 (2021).
45. N. J. Turner, *Ancient Pathways, Ancestral Knowledge: Ethnobotany and Ecological Wisdom of Indigenous Peoples of Northwestern North America*, (2 vols.), no. 74 of *Mcgill-Queen's Indigenous and Northern Studies*, J. Borrows, S. Carter, A. J. Ray, Eds. (McGill-Queen's Univ. Press, 2014).
46. J. T. Forrest, P. Kane, J. R. Harper, *West. Hist. Q.* **3**, 79–81 (1972).
47. J. G. Swan, *The Indians of Cape Flattery: At the Entrance to the Strait of Fuca, Washington Territory* (Smithsonian Institution, 1868).
48. C. H. Smith, *The Natural History of Dogs: Canidae or Genus Canis of Authors: Including Also the Genera Hyaena and Proteles* (W.H. Lizars, 1839).
49. C. Stantis, github/stantis/PNW-dogs-isotopes, Version v1.0, Zenodo (2023); https://10.5281/zenodo.10247167.

ACKNOWLEDGMENTS

We wish to express our deep gratitude to the Honorable S. Point, Grand Chief, and to G. Point of the Stó:lō Nation for giving us permission and encouragement for this research. Thanks to C. Wellman for her role in rediscovering Mutton, assistance with history of the area, and photographs. We raise our hands in thanks to all people within the Coast Salish communities who have graciously shared their time and knowledge to realize this project, specifically Xweliqwiya. R. Point-Bolton (Stó:lō Nation); D. Morsette (Suquamish/

Shxwhá:y Village); E. Kwulasultun White-Hill (Snuneymuxw First Nation); Sulqwan P. Williams (Cowichan); V. Snu'Meethia Elliott (Snuneymuxw); T. Sesemiya Williams (Skwxwú7mesh Úxwumixw/Squamish Nation); A. Fritz, Norris family (Lyacksun); T. Jones (Tulalip); T. Hohn (Puyallup); and q̓wa̓ləməmu N. Bob (Lummi). Interviews were carried out under Institutional Review Board and Research Ethics Board approvals from the Smithsonian Institution (Human Subjects Protocol no. HS220007) and Vancouver Island University (no. 101410), with informed consent including explicit opt-in permissions to reprint quotations with personal attribution. Computations performed for this paper were conducted on the Smithsonian High Performance Computing Cluster, Smithsonian Institution (https://doi.org/10.25572/SIHPC), and the Leibniz Supercomputing Centre (LRZ). Portions of the laboratory work were conducted in and with the support of the Laboratories of Analytical Biology (LAB) facilities of the National Museum of Natural History. Thanks to T. Gilbert for funding the processing/sequencing of AL3194, J. Ososky for specimen-handling assistance, and L. Orlando and S. Harding for providing helpful comments on the manuscript. **Funding:** Research was supported by Smithsonian Institution funds to L.K., A.T.L., H.-L.L., and C.S. were supported by Smithsonian postdoctoral fellowships. Funding for stable isotope analysis was provided by Smithsonian Museum Conservation Institute federal and trust funds. P.S. was supported by EMBO, the Vallee Foundation, the European Research Council (grant no. 852558), the Wellcome Trust (217223/Z/19/Z), and Francis Crick Institute core funding (FC001595) from Cancer Research UK, the Medical Research Council, and the Wellcome Trust. V.G. was supported by an SSHRC-IG. **Author contributions:** Conceptualization: A.T.L., L.H.-K., and L.K. Methodology: A.T.L., L.K., H.-L.L., L.H.-K., S.G.A., C.S., C.A.M.F., and K.C. Investigation: A.T.L., L.K., C.S., S.G.A., H.-L.L., M.T.R.H., L.H.-K., J.H., I.M., G.K., T.R.F., M.-H.S.S., S.G., L.F., A.B., A.C., A.H., and S.C. Formal analysis: A.T.L., L.K., C.S., C.A.M.F., S.G.A., D.W.G.S., and A.H. Visualization: A.T.L., L.K., C.S., K.C., M.H., G.K., and I.M. Resources: L.K., M.T.R.H., V.G., B.N.S., I.M., and E.A.O. Funding acquisition: L.K., P.S., and L.D. Supervision: L.K. and L.H.-.K. Writing – original draft: A.T.L., L.K., and L.H.-.K. Writing – review and editing: all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Genomic sequencing data for Mutton, SB dog, the Port au Choix dog (AL3194), and ALAS_015 are available for noncommercial use through NCBI SRA Project accession no. PRJNA1005336 and BioSample accession nos. SAMN36985984 to SAMN36985987. The SRA Project accession no. for the modern coyote from Wyoming is PRJNA734649. Stable isotope data are available (*49*). All other public genomic data sources are provided in data S1. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by The Wellcome Trust (217223/Z/19/Z), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi6549
Materials and Methods
Figs. S1 to S19
Tables S1 and S2
References (*50–161*)
MDAR Reproducibility Checklist
Data S1 to S5

Submitted 12 May 2023; accepted 25 October 2023
10.1126/science.adi6549

ATMOSPHERES

Iodine oxoacids enhance nucleation of sulfuric acid particles in the atmosphere

Xu-Cheng He[1,2,3]*, Mario Simon[4], Siddharth Iyer[5], Hong-Bin Xie[6]*, Birte Rörup[1], Jiali Shen[1], Henning Finkenzeller[7,8], Dominik Stolzenburg[1,9], Rongjie Zhang[6], Andrea Baccarini[10,11], Yee Jun Tham[1,12], Mingyi Wang[13], Stavros Amanatidis[13], Ana A. Piedehierro[3], Antonio Amorim[14], Rima Baalbaki[1], Zoé Brasseur[1], Lucía Caudillo[4], Biwu Chu[1,15], Lubna Dada[1,10], Jonathan Duplissy[1,16], Imad El Haddad[10], Richard C. Flagan[13], Manuel Granzin[4], Armin Hansel[17], Martin Heinritzi[4], Victoria Hofbauer[2], Tuija Jokinen[1,18], Deniz Kemppainen[1], Weimeng Kong[13], Jordan Krechmer[19], Andreas Kürten[4], Houssni Lamkaddam[10], Brandon Lopez[2,20], Fangfang Ma[6], Naser G. A. Mahfouz[2], Vladimir Makhmutov[21,22], Hanna E. Manninen[23], Guillaume Marie[4], Ruby Marten[10], Dario Massabò[24], Roy L. Mauldin[25,26], Bernhard Mentler[17], Antti Onnela[23], Tuukka Petäjä[1], Joschka Pfeifer[23], Maxim Philippov[21], Ananth Ranjithkumar[27], Matti P. Rissanen[5,28], Siegfried Schobesberger[29], Wiebke Scholz[17], Benjamin Schulze[13], Mihnea Surdu[10], Roseline C. Thakur[1], António Tomé[30], Andrea C. Wagner[4], Dongyu Wang[10], Yonghong Wang[1,15], Stefan K. Weber[23,4], André Welti[3], Paul M. Winkler[31], Marcel Zauner-Wieczorek[4], Urs Baltensperger[10], Joachim Curtius[4], Theo Kurtén[28], Douglas R. Worsnop[1,19], Rainer Volkamer[7,8], Katrianne Lehtipalo[1,3], Jasper Kirkby[23,4], Neil M. Donahue[2,20,25,32], Mikko Sipilä[1]*, Markku Kulmala[1,16,33,34]*

The main nucleating vapor in the atmosphere is thought to be sulfuric acid (H_2SO_4), stabilized by ammonia (NH_3). However, in marine and polar regions, NH_3 is generally low, and H_2SO_4 is frequently found together with iodine oxoacids [HIO_x, i.e., iodic acid (HIO_3) and iodous acid (HIO_2)]. In experiments performed with the CERN CLOUD (Cosmics Leaving OUtdoor Droplets) chamber, we investigated the interplay of H_2SO_4 and HIO_x during atmospheric particle nucleation. We found that HIO_x greatly enhances H_2SO_4(-NH_3) nucleation through two different interactions. First, HIO_3 strongly binds with H_2SO_4 in charged clusters so they drive particle nucleation synergistically. Second, HIO_2 substitutes for NH_3, forming strongly bound H_2SO_4-HIO_2 acid-base pairs in molecular clusters. Global observations imply that HIO_x is enhancing H_2SO_4(-NH_3) nucleation rates 10- to 10,000-fold in marine and polar regions.

Aerosols influence climate by acting as cloud condensation nuclei (CCN) and by scattering solar radiation. Secondary aerosol and CCN formation continue to be two of the largest uncertainties hindering accurate projection of climate change (*1*). Only a few types of vapors in the atmosphere can nucleate to form new aerosol particles, which can further grow to CCN sizes. Sulfuric acid (H_2SO_4) is considered to be the primary vapor (*2*) driving particle formation in the atmosphere of both polluted environments (*3*, *4*) and pristine environments (*5*–*7*). However, as H_2SO_4-H_2O binary nucleation is slow, stabilizing vapors, such as ammonia (NH_3), amines, and oxidized organics, are generally needed to explain observed particle formation rates (*3*–*11*).

In terms of radiative balance, marine clouds, especially low-level marine stratocumulus (*12*), are key players because they have strong longwave emission and efficiently reflect solar radiation back to space. As marine cloud formation is often limited by low CCN number concentrations, it is important to reach a comprehensive understanding of new particle formation in marine environments. New particle and subsequent CCN formation in marine regions is presently thought to be driven by H_2SO_4, methanesulfonic acid (MSA) (*8*, *13*), aided by NH_3 (*5*, *14*). However, a recent global survey of aerosol acidity suggests that global models substantially overestimate NH_3 concentrations; in particular, the polar atmosphere and high altitudes are characterized by low NH_3 concentrations (*15*). Assuming solely H_2SO_4 nucleation, advanced Earth system models struggle to reproduce aerosol number concentrations measured by aircraft (*16*), leading to low confidence for estimates of aerosol radiative forcing. Iodine-driven nucleation (*17*–*21*) has not yet been incorporated into Earth system models; iodine oxoacids (HIO_x, x = 2 to 3 in this study) can drive rapid particle formation under low NH_3 conditions, and they may play an important role in polar, marine, and free tropospheric particle formation.

In the marine atmosphere, iodine and sulfur precursors emitted from the ocean surface lead to the formation of both H_2SO_4 and HIO_x (*22*). HIO_x has generally been observed at concentrations similar to or lower than H_2SO_4 (*6*, *18*, *21*, *23*). Despite the higher nucleation potential of HIO_x compared with H_2SO_4 (*18*), iodine-driven new particle formation has hitherto been considered important only in regions with considerably higher concentrations of iodic acid (HIO_3) than of H_2SO_4, such as coastal zones and specific regions in the Arctic (*17*, *18*, *20*, *21*, *24*, *25*). However, new particle formation from the mixed chemical system HIO_x-H_2SO_4(-NH_3) has not been reported so far.

Particle formation experiments in CLOUD

Here we report laboratory experiments performed in the CERN CLOUD (Cosmics Leaving OUtdoor Droplets) chamber (*5*) (see methods in the supplementary materials for details) between September 2018 and December 2019 under conditions relevant for marine and polar environments. We performed particle formation experiments using HIO_x-H_2SO_4(-NH_3) vapors produced from the following precursors: molecular iodine (I_2), sulfur dioxide (SO_2), ammonia (NH_3), ozone (O_3), and water vapor (H_2O).

[1]Institute for Atmospheric and Earth System Research/Physics, Faculty of Science, University of Helsinki, 00014 Helsinki, Finland. [2]Center for Atmospheric Particle Studies, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [3]Finnish Meteorological Institute, 00560 Helsinki, Finland. [4]Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany. [5]Aerosol Physics Laboratory, Faculty of Engineering and Natural Sciences, Tampere University, 33720 Tampere, Finland. [6]Key Laboratory of Industrial Ecology and Environmental Engineering (Ministry of Education), School of Environmental Science and Technology, Dalian University of Technology, 116024 Dalian, China. [7]Department of Chemistry, University of Colorado Boulder, Boulder, CO 80309, USA. [8]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80309, USA. [9]Institute for Materials Chemistry, TU Wien, 1060 Vienna, Austria. [10]Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, CH-5232 Villigen, Switzerland. [11]Laboratory of Atmospheric Processes and their Impact, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. [12]School of Marine Sciences, Sun Yat-sen University, 519082 Zhuhai, China. [13]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [14]CENTRA and Faculdade de Ciências da Universidade de Lisboa, 1749-016 Lisboa, Portugal. [15]Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, 100084 Beijing, China. [16]Helsinki Institute of Physics, University of Helsinki, 00014 Helsinki, Finland. [17]Institute of Ion Physics and Applied Physics, University of Innsbruck, 6020 Innsbruck, Austria. [18]Climate and Atmosphere Research Centre (CARE-C), The Cyprus Institute, 1645 Nicosia, Cyprus. [19]Aerodyne Research, Inc., Billerica, MA 01821, USA. [20]Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [21]P. N. Lebedev Physical Institute of the Russian Academy of Sciences, 119991 Moscow, Russia. [22]Moscow Institute of Physics and Technology (National Research University),141701 Moscow, Russian Federation. [23]CERN, the European Organization for Nuclear Research, CH-1211 Geneva, Switzerland. [24]Department of Physics, University of Genoa, 16146 Genoa, Italy. [25]Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [26]Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, CO 80309, USA. [27]Natural Environment Research Council, British Antarctic Survey, CB3 0ET Cambridge, UK. [28]Department of Chemistry, University of Helsinki, 00014 Helsinki, Finland. [29]Department of Applied Physics, University of Eastern Finland, 70211 Kuopio, Finland. [30]Instituto Dom Luiz (IDL)–Universidade da Beira Interior, 6201-001 Covilhã, Portugal. [31]Faculty of Physics, University of Vienna, 1090 Wien, Austria. [32]Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [33]Joint International Research Laboratory of Atmospheric and Earth System Sciences, School of Atmospheric Sciences, Nanjing University, 210023 Nanjing, China. [34]Aerosol and Haze Laboratory, Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, 100029 Beijing, China.
*Corresponding author. Email: xucheng.he@helsinki.fi (X.-C.H.); markku.kulmala@helsinki.fi (M.K.); mikko.sipila@helsinki.fi (M.Sip.); hbxie@dlut.edu.cn (H.-B.X.)

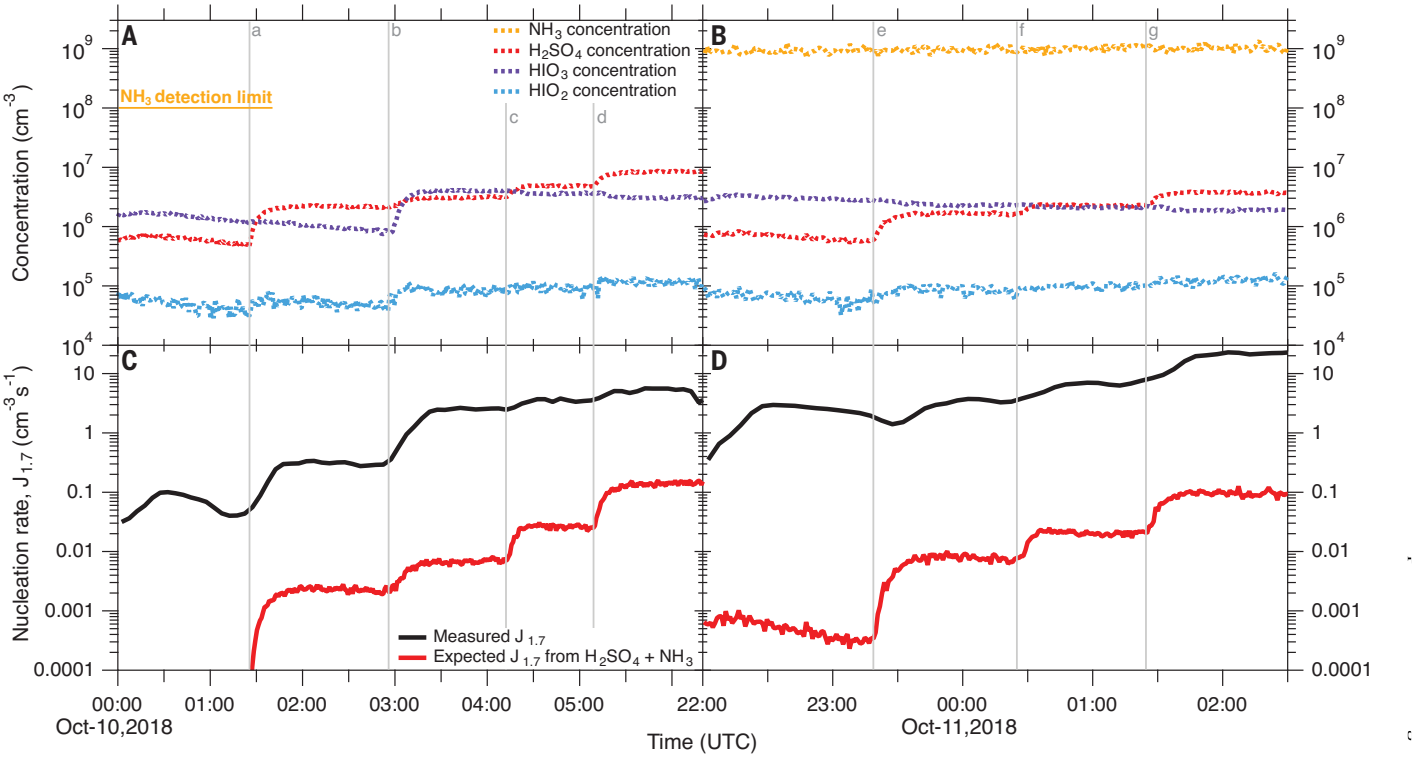


Fig. 1. New particle formation from HIO_x-H_2SO_4 and HIO_x-H_2SO_4-NH_3 at −10°C. (**A** and **B**) vapor concentrations and (**C** and **D**) nucleation rates. Solid black lines show the measured nucleation rates at 1.7 nm and solid red lines present predicted $J_{1.7}$ from H_2SO_4-NH_3 nucleation alone (*14*). Dashed lines represent vapor concentrations, and vertical gray bars show experimental stages. The experiments show that the rapid nucleation rates cannot be explained by the H_2SO_4-NH_3 mechanism alone. HIO_x significantly enhances H_2SO_4-NH_3 nucleation at comparable HIO_3 and H_2SO_4 concentrations. The NH_3 concentration in (A) is below the detection limit of the H_3O^+-CIMS (~4 pptv). An NH_3 concentration of 4 pptv is used to conservatively estimate the H_2SO_4-NH_3 nucleation rates in (C). The experimental conditions are 41.1 parts per billion by volume (ppbv) O_3, 63.5% relative humidity (RH), 2.3 ppbv SO_2, and 17.4 pptv I_2 [(A) and (C)]; and 40.8 ppbv O_3, 62.3% RH, 1.6 ppbv SO_2, and 67.2 pptv I_2 [(B) and (D)]. Stages a, c, d, e, f, and g enhanced the UVH light intensity (higher OH production rates), and stage b increased the green light intensity (higher I_2 photolysis rate).

To investigate possible synergies in HIO_x-H_2SO_4(-NH_3) nucleation, green and ultraviolet light sources were used to drive photochemical production of HIO_x and H_2SO_4 from I_2 and SO_2. An example experiment at −10°C is shown in Fig. 1 and fig. S1, and at 10°C in fig. S2. Experiments were first performed without any added NH_3 [<4 parts per trillion by volume (pptv) contaminant level]; these are shown in the left-hand panels of Fig. 1 and figs. S1 and S2. A second set of experiments were performed with NH_3 added to the chamber (right-hand panels of Fig. 1 and figs. S1 and S2). At both temperatures, a significantly higher nucleation rate at 1.7 nm, $J_{1.7}$, is observed in the presence of HIO_x than the $J_{1.7}$ expected from H_2SO_4-NH_3 nucleation (*5*, *14*), both without and with added NH_3.

In Fig. 2, we present $J_{1.7}$ for the HIO_x-H_2SO_4 system (hollow markers) and the HIO_x-H_2SO_4-NH_3 system (filled markers) at 10°C (circles) and −10°C (squares). The concentration ranges of HIO_x and H_2SO_4 closely match ambient values, spanning from $<10^6$ cm^{-3} to nearly 10^8 cm^{-3} (*6*, *17*, *18*, *20*, *21*, *23*). We show the measured $J_{1.7}$ for these mixed systems for various possible drivers: H_2SO_4 (Fig. 2A), HIO_3 + H_2SO_4 (Fig. 2B), and (HIO_3 + H_2SO_4) × HIO_2 (Fig. 2C) (HIO_2, iodous acid). The data at both temperatures become progressively less scattered when plotted against these variables, as well as more consistent with parameterizations (*14*, *18*). The H_2SO_4-NH_3 mechanism cannot predict the nucleation rates, even when the HIO_x concentration is much lower than that of H_2SO_4 (Fig. 2A). For instance, $J_{1.7}$ at 10°C from HIO_x-H_2SO_4 with NH_3 < 4 pptv (Fig. 2A, hollow circles) is roughly 60 times faster than $J_{1.7}$ from H_2SO_4 with NH_3 at 4 pptv; this is as fast as nucleation from H_2SO_4 with NH_3 at 500 pptv. Therefore, sub-pptv levels of HIO_x are as effective at stabilizing H_2SO_4 as 500 pptv of NH_3. Hence, HIO_x may replace NH_3 as a nucleation driver in pristine marine and polar environments, where NH_3 concentrations are typically below a few tens of parts per trillion by volume or lower (*26*, *27*).

Figure 2B shows the observed $J_{1.7}$ versus total acid concentration (HIO_3 + H_2SO_4) and compares these rates to the values predicted by the H_2SO_4(-NH_3) parameterizations (*14*), applying (HIO_3 + H_2SO_4) as H_2SO_4. The $J_{1.7}$ of the HIO_x-H_2SO_4 system without added NH_3 (hollow markers) remains higher than the prediction for H_2SO_4(-NH_3) nucleation. This indicates that HIO_x contributes more prominently to nucleation than by simply increasing the acid concentration. Moreover, the relatively mild sensitivity to NH_3 suggests that the base stabilization comes from another source. This is supported by Fig. 2C, which indicates that HIO_2 is effectively providing base stabilization in the molecular clusters. To further investigate the underlying mechanisms, we studied the molecular composition of nucleating particles under neutral (ion-free) and charged (ion-induced) conditions, as described below.

HIO_2 accelerates neutral nucleation

To measure neutral clusters, we used a nitrate chemical ionization mass spectrometer (nitrate-CIMS). The concentrations of monomers HIO_3, H_2SO_4, and HIO_2 are presented in Fig. 3A, together with four product dimers in Fig. 3B. Although the HIO_2 concentration was one to two orders of magnitude lower than that of HIO_3 or H_2SO_4, the most prominent dimers, HIO_3-HIO_2 and H_2SO_4-HIO_2, both contain HIO_2.

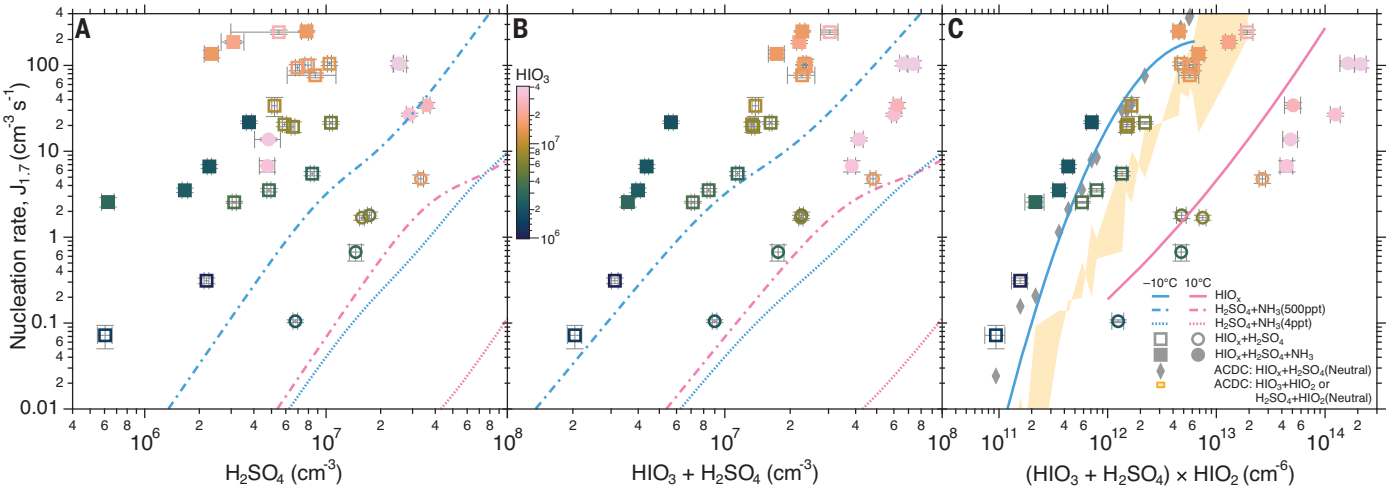


Fig. 2. Nucleation rates of HIO_x-H_2SO_4(-NH_3) systems. Nucleation rates at 1.7 nm versus (**A**) H_2SO_4, (**B**) HIO_3 + H_2SO_4, and (**C**) (HIO_3 + H_2SO_4) × HIO_2 at +10° and −10°C. All data points and lines show experiments carried out at galactic cosmic ray ionization conditions, except for the atmospheric cluster dynamics code (ACDC) simulations in (C) (orange band and filled diamonds), which represent the theoretical prediction for the neutral nucleation rates (see methods). The color bar represents HIO_3 concentration (per cubic centimeter). H_2SO_4-NH_3 mechanism fails to predict the overall nucleation rates, even with HIO_x is much lower than H_2SO_4. The $J_{1.7}$ from experiments with high H_2SO_4 is also higher than that predicted by pure iodine oxoacids (*18*). The nucleation rates become less spread when plotted against (HIO_3 + H_2SO_4) × HIO_2, as well as more consistent with parameterizations and ACDC predictions. The results show that HIO_3 and HIO_2 have to be considered together with H_2SO_4 to predict the nucleation rates in this multicomponent system. H_2SO_4-NH_3 nucleation rates (dotted and dash-dotted lines) are calculated following Dunne *et al*. (*14*), whereas HIO_x nucleation rates (solid lines) are calculated on the basis of $J_{1.7}$, HIO_3, and recalculated HIO_2 from He *et al*. (*18*), applying HIO_3 × HIO_2 as (HIO_3 + H_2SO_4) × HIO_2, to guide the eye. The experimental conditions for HIO_x-H_2SO_4(-NH_3) experiments are 38.4 to 53.2 ppbv O_3, 41.9 to 75.3% RH, 0.6 to 11.2 ppbv SO_2, and 10.0 to 57.7 pptv I_2. The NH_3 concentrations for the filled squares and filled circles range from 30 to 42 pptv and from 176 to 261 pptv, respectively. The error bars show one standard deviation during the data selection periods. Overall systematic scale errors on the HIO_3 concentrations of −33% and +50% and on the nucleation rates of a factor of 10 are not shown on the data points.

Despite HIO_3-HIO_2 clusters having been reported before (*18*, *28*), we believe this is the first observation of H_2SO_4-HIO_2 dimers.

While HIO_2 enables H_2SO_4-HIO_2 dimer formation, its role in larger clusters is not clear. We address this with a combination of quantum chemical calculations and cluster dynamics modeling (*29*). We optimized the geometries of H_2SO_4-HIO_2, HIO_3-HIO_2, and H_2SO_4-HIO_3-HIO_2 clusters and calculated their formation free energies and evaporation rates (fig. S3). Clusters containing HIO_2 are the most stable and, moreover, show an exceptionally wide range of stable combinations of molecules. The cluster geometries suggest that HIO_2 enhances H_2SO_4 neutral nucleation in the same way as it does for HIO_3 neutral nucleation (*18*). Specifically, HIO_2 accepts the proton donated either by H_2SO_4 or HIO_3, thereby functioning as a base. Furthermore, HIO_2 forms strong halogen bonds with H_2SO_4 and HIO_3, further enhancing the cluster stability. Clusters including HIO_2 are even more stable than H_2SO_4-DMA (dimethyl amine) clusters (fig. S3), which is known to cluster at the collision limit for sulfuric acid with only 4 pptv DMA (*3*). However, the predicted neutral nucleation rates for the H_2SO_4-HIO_2 and HIO_3-HIO_2 systems still underestimate our measured nucleation rates [galactic cosmic ray (GCR) conditions, the sum of neutral and ion-induced channels] at −10°C (Fig. 2C, orange band). On the other hand, the predicted HIO_x-H_2SO_4 neutral nucleation rates approximately agree with CLOUD observations (Fig. 2C, squares and diamonds). The consistency between theoretical predictions and the CLOUD measurements at −10°C suggests that neutral nucleation dominates at this temperature, which is also indicated by the fact that the nucleation rates far exceed the ion-pair production rate limit (2 to 10 cm^{-3} s^{-1}). Additionally, this suggests that the controlling mechanism is indeed a synergy of three molecules (HIO_3, H_2SO_4, and HIO_2) and not simply the combined neutral nucleation of any two molecules. Given that HIO_2 behaves as a base, we show in Fig. 2C our observed $J_{1.7}$ versus (HIO_3 + H_2SO_4) × HIO_2. This expression is proportional to the formation rate of the dimer (H_2SO_4-HIO_2 or HIO_3-HIO_2), which represents the initial nucleating cluster. We find that the HIO_x-H_2SO_4(-NH_3) nucleation rates fall near the prediction from HIO_x nucleation ($J_{1.7}$ versus HIO_3 × HIO_2; H_2SO_4 is absent in pure iodine oxoacid nucleation, but it is added to the HIO_3 concentration given its identical role) (*18*), implying that HIO_2 indeed plays the key role as stabilizer both for HIO_3 and H_2SO_4 and that NH_3 plays a minor role.

While the formation mechanism for HIO_3 has recently been established (*22*), the pathway for HIO_2 formation remains uncertain. A quantum chemical study provided a potential energy surface describing formation of HIO_2 from iodooxy hypoiodite, I_2O_2 + H_2O (*30*). We have extended this study with high-level quantum chemical calculations and provide a revised potential energy surface in fig. S4A. We also present a potential new pathway for HIO_2 formation from iodine dioxide (OIO) and the hydroperoxyl radical (HO_2) (fig. S4B). Our calculations show that both the singlet and triplet channels are exothermic. Further studies are needed to quantify the relative importance in the atmosphere of these two channels.

Because complex reactions are involved in the formation of HIO_3 and HIO_2, it is important to confirm that the HIO_3:HIO_2 ratio in the CLOUD chamber matches ambient conditions. Figure S5 shows that both the ratio and absolute concentrations of HIO_3 and HIO_2 fall within the range measured at Mace Head (*17*) and Ny-Ålesund (*21*), confirming that the results from our study are relevant to the atmosphere.

HIO_3 enhances ion-induced nucleation

Ions can stabilize embryonic molecular clusters, leading to ion-induced nucleation (IIN) (*5*, *18*, *19*, *31*, *32*). To investigate the influence of ions on HIO_x-H_2SO_4 nucleation, we increased the ionization rate in the chamber in three steps at 10°C: (i) neutral (ion-free), (ii) GCR ionization (~1000 ion pairs cm^{-3}), and (iii) beam-enhanced ionization (~6000 ion pairs cm^{-3}) (fig. S6). Compared with neutral conditions, $J_{1.7}$ at GCR ionization rates is enhanced by

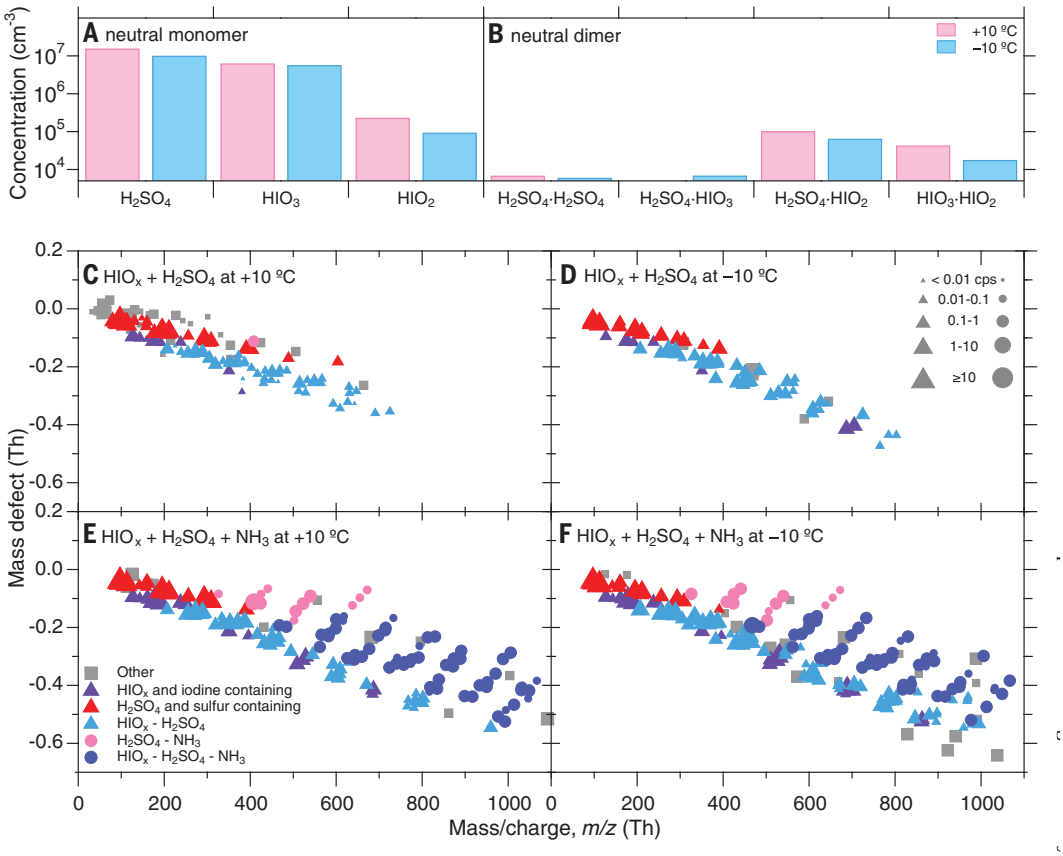


Fig. 3. Neutral and charged cluster composition during HIO_x-H_2SO_4-NH_3 nucleation. Background-subtracted neutral monomer (**A**) and dimer (**B**) concentrations in HIO_x-H_2SO_4 nucleation events at +10°C (pink bars) and −10°C (cyan bars). (**C** and **D**) Negatively charged cluster compositions of HIO_x-H_2SO_4 nucleation at +10° and −10°C, respectively. (**E** and **F**) Negatively charged cluster compositions of HIO_x-H_2SO_4-NH_3 nucleation at +10° and −10°C, respectively. As indicated in (B), the dominant neutral dimers are H_2SO_4-HIO_2 and HIO_3-HIO_2 clusters—despite very low HIO_2 concentrations—which represent the initial molecular clusters during neutral nucleation. Ion-induced nucleation is dominated by charged HIO_3-H_2SO_4 [(C) and (D)] or HIO_3-H_2SO_4(-NH_3) [(E) and (F)] cluster formation processes. HIO_3-NH_3 clusters are not detected, which suggests that NH_3 has a negligible effect on ion-induced HIO_3 cluster formation. The marker size is shown in the legend (cps, ion counts per second). The experimental conditions are 38.5 to 43.9 ppbv O_3, 61.6 to 75.2% RH, 0.7 to 11.0 ppbv SO_2, and 14.4 to 44.5 pptv I_2.

~50 times at 2×10^7 cm^{-3} H_2SO_4 and 5×10^6 cm^{-3} HIO_3. As with HIO_x, ion-induced HIO_x-H_2SO_4 nucleation only occurs with negative ions (compare fig. S6A and fig. S6B). Interestingly, six times larger ion concentrations formed by the pion beam only enhance $J_{1.7}$ by a factor of two. This is likely because the increased ion-ion recombination rate, and hence the shorter charge lifetime, neutralizes some clusters before they have become stable against evaporation when neutral. When NH_3 is added to the HIO_x-H_2SO_4 system, it initiates positive IIN that is as strong as negative IIN at −10°C (fig. S1D). Adding NH_3 approximately doubles the overall $J_{1.7}$.

To measure the molecular composition of charged clusters, we used an atmospheric pressure interface time-of-flight mass spectrometer. For HIO_x-H_2SO_4 IIN without NH_3 injection (Fig. 3, C and D), we observe a series of charged clusters with the empirical formula $(HIO_3)_n$-$(H_2SO_4)_m$-HSO_4^- (cyan triangles), which indicate synergistic IIN of HIO_3 and H_2SO_4. We identify these clusters as (n+m+1)-mer (which include the ion; fig. S7). At 10°C, monomers, dimers, and trimers consist primarily of H_2SO_4, whereas HIO_3 appears in clusters starting from the tetramers and becomes equal to the H_2SO_4 mole fraction already in the hexamers. At −10°C, HIO_3 appears in the dimers and becomes equal to the H_2SO_4 starting with the tetramers.

In these experiments, the HIO_3:H_2SO_4 ratio in the gas phase is between 0.3 and 1.4, and the molar ratios of I:S in the larger clusters tend toward 1:1. We also know that pure ion-induced HIO_3 nucleation proceeds at the collision limit (*18*, *19*) but that ion-induced H_2SO_4 nucleation is slower than the collision limit (*5*). We therefore conclude that H_2SO_4 condensation is enhanced by HIO_3 for a cluster stoichiometry up to 1:1, beyond which the net rate of H_2SO_4 condensation slows, while HIO_3 condensation is limited by the collision rate under our experimental conditions.

We performed additional experiments in which NH_3 was added to the HIO_x-H_2SO_4 system. Notably, none of the charged pure iodine (i.e., H_2SO_4-free) clusters contained NH_3, which indicates a negligible role of NH_3 in ion-induced HIO_3 nucleation. This was independently confirmed by raising NH_3 from the background level (<4 pptv) to 100 pptv in an iodine oxoacid nucleation experiment without H_2SO_4 (fig. S8). The measured nucleation rate at 1.7 nm remained constant throughout the experiment, indicating that HIO_3(-HIO_2) nucleation is unaffected by NH_3.

On the other hand, we found a set of clusters with the composition $(H_2SO_4)_n$-$(NH_3)_m$-HSO_4^- and $(HIO_3)_n$-$(H_2SO_4)_m$-$(NH_3)_j$-HSO_4^- in the mass spectra of charged clusters (Fig. 3, E and F), similar to the clusters reported near the coast of Antarctica (*6*). We found that NH_3 is only present in charged tetramers and above, consistent with its behavior in H_2SO_4-NH_3 IIN (*5*). The iodine and sulfur molar fraction distributions remained unchanged after adding NH_3 to the system, likely because the HIO_3-H_2SO_4 negative IIN had already reached the collision limit (fig. S7). The presence of NH_3 only converted some of the $(HIO_3)_n$-$(H_2SO_4)_m$-HSO_4^- to $(HIO_3)_n$-$(H_2SO_4)_m$-$(NH_3)_j$-HSO_4^- ions and gives rise to positive IIN (fig. S1).

Particle growth

Since the atmospheric concentration of HIO_2 is less than one-tenth that of HIO_3, its role in particle growth is minor (*18*). To evaluate the role of HIO_3 and H_2SO_4 in particle growth, we compare in fig. S9A our measured growth rates between 1.8 and 3.2 nm ($GR_{1.8\text{-}3.2}$) with those calculated assuming condensation of H_2SO_4 and HIO_3 (*18*, *33*) at the collision limit. The good agreement indicates that H_2SO_4 and HIO_3 are the main condensing vapors driving particle growth (*18*, *33*) while other iodine species contribute little to particle mass. We show in fig. S9B the measured and predicted particle survival probability, $J_{2.5}/J_{1.7}$, which increases at faster growth rates and approaches unity above growth rates of ~10 nm hour^{-1} for the CLOUD chamber (2.2×10^{-3} s^{-1} wall loss rate). In the marine atmosphere, condensation

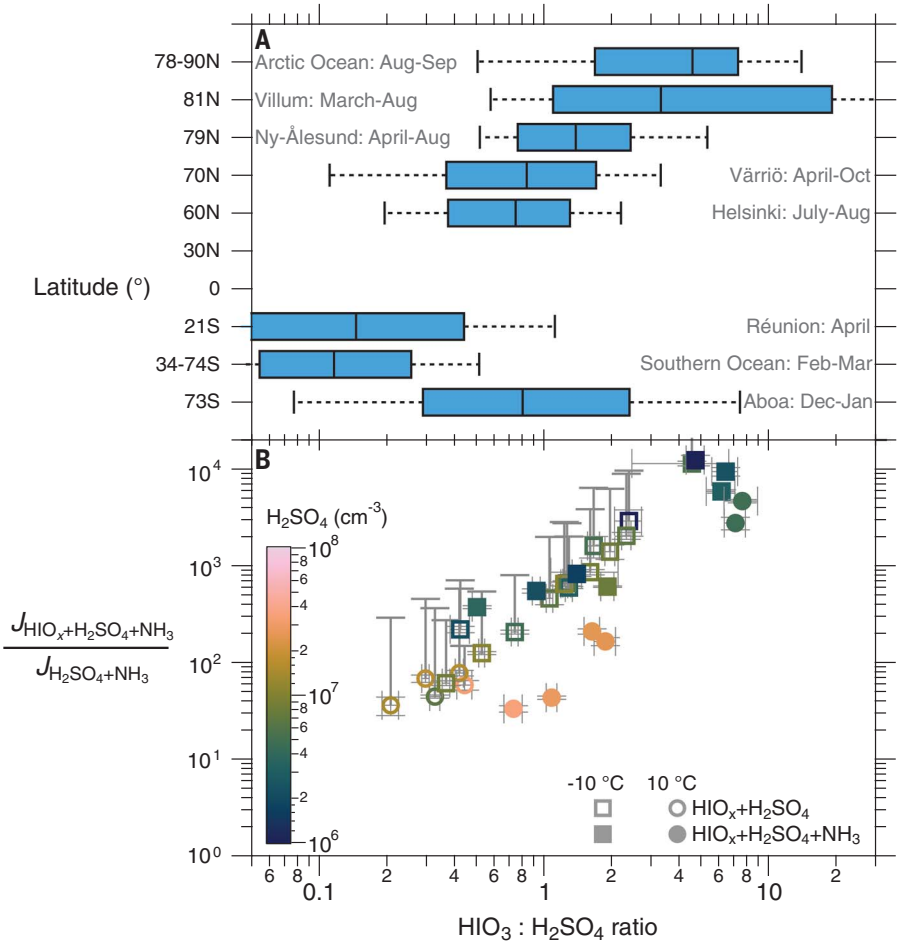


Fig. 4. HIO_x enhancement of H_2SO_4-NH_3 nucleation. (**A**) Box plot statistics of HIO_3:H_2SO_4 ratios measured around the globe, showing median values with 10 and 90% percentiles in the whiskers. The gray labels show the site name and months, with data covering >50% of the days. (**B**) Nucleation enhancement by HIO_x is calculated by dividing the measured $J_{HIO_x-H_2SO_4-NH_3}$ to predict $J_{H_2SO_4-NH_3}$ using CLOUD parameterizations (*14*). The median ratios at all sites are greater than 0.1, which infers at least a 10-fold nucleation rate enhancement by HIO_x. The enhancement is especially pronounced in polar regions where the HIO_3:H_2SO_4 ratio is consistently higher than 0.1. Thin symmetric error bars represent one standard deviation during the data selection periods. In the experiments without NH_3 injection (hollow markers), the NH_3 concentrations were below the instrument detection limit (4 pptv), which is adopted as a conservative estimate of $J_{H_2SO_4-NH_3}$. However, the actual NH_3 concentration is expected to be below 1 pptv, as all charged clusters are essentially NH_3 free (Fig. 3). The thick asymmetric error bars represent the systematic uncertainty assuming NH_3 equals 1 pptv. The NH_3 concentrations in experiments with NH_3 injection (filled markers) are well measured, and thus without asymmetric errors. The field observation sites are summarized in the methods.

of other compounds, such as MSA and oxidized organic molecules, can also contribute to early particle growth, in addition to H_2SO_4 and HIO_3.

Climate implications

Atmospheric observations show that both iodine oxoacid and sulfuric acid–ammonia nucleation can be important particle sources in specific regions of the pristine boundary layer (*6*, *17*, *18*, *20*, *21*). So far, HIO_3 and HIO_2 have been thought to be important only in regions where they are more abundant than H_2SO_4. In polar and marine environments, it is currently thought that H_2SO_4-NH_3 constitutes the primary source of new particle formation, despite the perceived scarcity of NH_3 (*15*). This picture is challenged by our findings. Our data support the reverse: H_2SO_4-NH_3 nucleation plays a major role only when H_2SO_4 is substantially more abundant than HIO_3 and HIO_2. The role of HIO_x in atmospheric aerosol nucleation may have been overlooked, as studies could easily be deceived by relatively higher H_2SO_4 than HIO_x in parts of the pristine atmosphere.

To assess the atmospheric importance of HIO_x-H_2SO_4(-NH_3) nucleation, we calculated the $J_{1.7}$ enhancement factor [the ratio of $J_{1.7}$ from HIO_x-H_2SO_4(-NH_3) to that from H_2SO_4-NH_3] (*14*) as a function of the HIO_3:H_2SO_4 concentration ratio (Fig. 4B). The enhancement factors are large, ranging from 10 to 10^4 for atmospherically relevant HIO_3:H_2SO_4 ratios. Even when the HIO_3:H_2SO_4 ratio is 0.1, the enhancement factor is 10. Observations at marine and polar sites from the North Pole to Antarctica show median HIO_3:H_2SO_4 ratios larger than 0.1 (Fig. 4A), implying that synergistic HIO_x-H_2SO_4(-NH_3) nucleation may have global importance and yet has hitherto been overlooked. This conclusion is supported by our calculations of sulfuric acid nucleation enhanced by HIO_x, which are shown in fig. S10. At −10°C, which is representative of the marine free troposphere, fast nucleation rates of up to 10 cm^{-3} s^{-1} are estimated for ambient acid concentrations. The pronounced temperature dependence of HIO_x-H_2SO_4(-NH_3) nucleation that we find in our study may help explain why nucleation in the marine boundary is rarely observed, whereas nucleation is frequently found in the free troposphere or the upper marine boundary layer after passage of a cold front (*23*, *34*, *35*).

New particle formation from HIO_x-H_2SO_4 has notable implications for the future climate. Iodine oxoacids may enhance CCN and cloud formation in the Arctic (*20*), which would, in turn, affect both long- and shortwave radiative forcing at the surface (*36*). The absence of iodine oxoacid nucleation mechanisms in climate models may help explain why they systematically underestimate the CCN number concentration around the coast of Antarctica (*37*, *38*). Iodine has also been observed in both gas and particle phases in the polar and marine free troposphere and the upper troposphere and lower stratosphere (*39*, *40*). These regions are characterized by low temperatures and extremely low NH_3 concentrations (*15*), conditions that strongly favor HIO_x-H_2SO_4 or pure HIO_x nucleation over H_2SO_4-NH_3 nucleation. While global anthropogenic SO_2 emissions continue to fall as a result of emission policies, iodine emissions have tripled since the 1950s, and this trend continues (*41*, *42*). As a result, nucleation mechanisms involving iodine oxoacids are anticipated to become even more important in future. To sharpen the understanding of marine aerosol-cloud radiative forcing, it is important that representations of new particle formation in global climate models now include iodine oxoacids together with sulfuric acid.

REFERENCES AND NOTES

1. V. Masson-Delmotte *et al.*, Eds., *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2021).
2. V.-M. Kerminen *et al.*, *Environ. Res. Lett.* **13**, 103003 (2018).
3. J. Almeida *et al.*, *Nature* **502**, 359–363 (2013).
4. M. Xiao *et al.*, *Atmos. Chem. Phys.* **21**, 14275–14291 (2021).
5. J. Kirkby *et al.*, *Nature* **476**, 429–433 (2011).

6. T. Jokinen *et al*., *Sci. Adv.* **4**, eaat9744 (2018).
7. J. Brean *et al*., *Nat. Geosci.* **14**, 383–388 (2021).
8. S. M. Kreidenweis, J. H. Seinfeld, *Atmos. Environ.* **22**, 283–296 (1988).
9. P. H. McMurry, *J. Colloid Interface Sci.* **78**, 513–527 (1980).
10. C. Kuang, P. H. McMurry, A. V. McCormick, F. L. Eisele, *J. Geophys. Res.* **113**, D10209 (2008).
11. C. Yan *et al*., *Geophys. Res. Lett.* **48**, e2020GL091944 (2021).
12. R. Wood, *Mon. Weather Rev.* **140**, 2373–2423 (2012).
13. M. L. Dawson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18719–18724 (2012).
14. E. M. Dunne *et al*., *Science* **354**, 1119–1124 (2016).
15. B. A. Nault *et al*., *Commun. Earth Environ.* **2**, 93 (2021).
16. A. Ranjithkumar *et al*., *Atmos. Chem. Phys.* **21**, 4979–5014 (2021).
17. M. Sipilä *et al*., *Nature* **537**, 532–534 (2016).
18. X.-C. He *et al*., *Science* **371**, 589–595 (2021).
19. X.-C. He *et al*., *Aerosol Sci. Technol.* **55**, 231–242 (2021).
20. A. Baccarini *et al*., *Nat. Commun.* **11**, 4924 (2020).
21. L. J. Beck *et al*., *Geophys. Res. Lett.* **48**, e2020GL091334 (2021).
22. H. Finkenzeller *et al*., *Nat. Chem.* **15**, 129–135 (2023).
23. A. Baccarini *et al*., *J. Geophys. Res. Atmos.* **126**, e2021JD035126 (2021).
24. T. Hoffmann, C. D. O'Dowd, J. H. Seinfeld, *Geophys. Res. Lett.* **28**, 1949–1952 (2001).
25. C. D. O'Dowd *et al*., *Nature* **417**, 632–636 (2002).
26. G. R. Wentworth *et al*., *Atmos. Chem. Phys.* **16**, 1937–1953 (2016).
27. K. E. Altieri, K. A. M. Spence, S. Smith, *Geophys. Res. Lett.* **48**, e2020GL091963 (2021).
28. R. Zhang *et al*., *Environ. Sci. Technol.* **56**, 14166–14177 (2022).
29. M. J. McGrath *et al*., *Atmos. Chem. Phys.* **12**, 2345–2355 (2012).
30. J. C. Gómez Martín *et al*., *Nat. Commun.* **11**, 4521 (2020).
31. S.-H. Lee *et al*., *Science* **301**, 1886–1889 (2003).
32. E. R. Lovejoy, J. Curtius, K. D. Froyd, *J. Geophys. Res.* **109**, D0820 (2004).
33. D. Stolzenburg *et al*., *Atmos. Chem. Phys.* **20**, 7359–7372 (2020).
34. K. J. Sanchez *et al*., *Atmos. Chem. Phys.* **21**, 3427–3446 (2021).
35. G. Zheng *et al*., *Nat. Commun.* **12**, 527 (2021).
36. T. Mauritsen *et al*., *Atmos. Chem. Phys.* **11**, 165–173 (2011).
37. I. L. McCoy *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18998–19006 (2020).
38. L. A. Regayre *et al*., *Atmos. Chem. Phys.* **20**, 10063–10072 (2020).
39. R. Volkamer *et al*., *Atmos. Meas. Tech.* **8**, 2121–2148 (2015).
40. T. K. Koenig *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1860–1866 (2020).
41. C. A. Cuevas *et al*., *Nat. Commun.* **9**, 1452 (2018).
42. M. Legrand *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12136–12141 (2018).
43. X.-C. He *et al*., Iodine oxoacids enhance nucleation of sulfuric acid particles in the atmosphere: data resources, version 1, Zenodo (2023); https://doi.org/10.5281/zenodo.8344386.

ACKNOWLEDGMENTS

We thank CERN for supporting CLOUD with important technical and financial resources and for providing a particle beam from the CERN Proton Synchrotron. We thank the CSC IT Center for Science in Espoo, Finland, for computing time. **Funding:** This work was supported by Research Council of Finland ACCC Flagship nos. 337549 and 337552; Research Council of Finland professorship (302958); Research Council of Finland Centre of Excellence nos. 346371, 346372, and 346373; Research Council of Finland project nos. 335844, 334514, 345125, 325656, 316114, 314798, 325647, 353836, 341349, 346372, 296628, and 349659; European Research Council (ERC) project nos. 742206, 714621, 616075, and 101002728; Marie Skłodowska-Curie grant agreement nos. 316662, 701647, 895875, and 764991; National Natural Science Foundation of China grant nos. 42175118 and 22236004; the National Key Research and Development Program of China (2022YFC3701000, Task 1); Swiss National Science Foundation grant nos. 200021_169090, 200021_213071, 206021_198140, 200020_172602, and 20FI20_172622; US National Science Foundation grants nos. AGS-2132089, AGS-1801897, AGS-1801574, AGS-1801280, AGS-2027252, AGS-2215522, AGS-1602086, AGS-1801329, and AGS-2215527; German Federal Ministry of Education and Research (BMBF) project no. 01LK1601A; and Portuguese Science Foundation project CERN/FIS-COM/0028/2019. R.C.T. and M.K. thank the Jane and Aatos Erkko Foundation for funding. M.K. received funding from Prince Albert Foundation contract no. 2859. M.W. thanks the Schmidt Science Fellowship for support. X.-C.H. and M.K. thank the Jenny and Antti Wihuri Foundation for funding. **Author contributions:** X.-C.H., J.Kir., M.Sip., and M.K. planned the experiments. X.-C.H., M.Sim., B.R., J.S., H.F., D.S., A.B., Y.J.T., M.W., S.A., A.A.P., A.A., R.B., Z.B., L.C., B.C., L.D., J.D., I.E.H., R.C.F., M.G., M.H., V.H., W.K., J.Kre., A.K., H.L., B.L., N.G.A.M., V.M., H.E.M., G.M., R.M., D.M., R.L.M., B.M., A.O., T.P., J.P., M.P., M.P.R., S.S., W.S., B.S., M.Sur., A.T., A.C.W., D.W., Y.W., S.K.W., A.W., P.M.W., M.Z.-W., J.Kir., U.B., K.L., J.C., R.V., M.Sip., and M.K. prepared the CLOUD facility or measuring instruments. X.-C.H., M.Sim., S.I., B.R., J.S., H.F., D.S., A.B., Y.J.T., A.A., R.B., L.C., L.D., J.D., I.E.H., M.G., M.H., V.H., T.J., D.K., W.K., J.Kre., H.L., B.L., N.G.A.M., V.M., H.E.M., G.M., R.M., D.M., R.LM., B.M., J.P., A.R., W.S., R.C.T., A.C.W., D.W., S.K.W., M.Z.-W., J.Kir., J.C., and R.V. collected the data. X.-C.H., S.I., H.-B.X., H.F., D.S., R.Z., Y.J.T., L.C., L.D., R.C.F., F.M., W.S., R.C.T., S.K.W., P.M.W., T.K. and D.R.W. analyzed the data. X.-C.H., N.M.D. and J.Kir. wrote the manuscript with contributions from S.I., H.-B.X., B.R. and R.Z. X.-C.H., M.Sim., S.I., H.-B.X., B.R., J.S., H.F., D.S., R.Z., A.B., L.D., J.D., I.E.H., R.C.F., A.H., H.L., N.G.A.M., T.P., S.S., W.S., R.C.T., A.C.W., J.Kir., U.B., K.L., J.C., T.K., D.R.W., R.V., N.M.D., M.Sip. and M.K. commented on and edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Data for all figures in the main text and supplementary materials are available at the Zenodo repository (*43*). Correspondence and additional requests for materials should be addressed to X.-C.H. (institutional email: xucheng.he@helsinki.fi; permanent email: xuchenghe93@gmail.com). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh2526
Materials and Methods
Figs. S1 to S10
References (*44*–*76*)

Submitted 20 February 2023; accepted 13 November 2023
10.1126/science.adh2526

PLANT OPTOGENETICS

Light-gated channelrhodopsin sparks proton-induced calcium release in guard cells

Shouguang Huang[1]*, Like Shen[1,2], M. Rob G. Roelfsema[1], Dirk Becker[1], Rainer Hedrich[1]*

Although there has been long-standing recognition that stimuli-induced cytosolic pH alterations coincide with changes in calcium ion (Ca^{2+}) levels, the interdependence between protons (H^+) and Ca^{2+} remains poorly understood. We addressed this topic using the light-gated channelrhodopsin *Hc*KCR2 from the pseudofungus *Hyphochytrium catenoides*, which operates as a H^+ conductive, Ca^{2+} impermeable ion channel on the plasma membrane of plant cells. Light activation of *Hc*KCR2 in *Arabidopsis* guard cells evokes a transient cytoplasmic acidification that sparks Ca^{2+} release from the endoplasmic reticulum. A H^+-induced cytosolic Ca^{2+} signal results in membrane depolarization through the activation of Ca^{2+}-dependent SLAC1/SLAH3 anion channels, which enabled us to remotely control stomatal movement. Our study suggests a H^+-induced Ca^{2+} release mechanism in plant cells and establishes *Hc*KCR2 as a tool to dissect the molecular basis of plant intracellular pH and Ca^{2+} signaling.

Eukaryotes move protons across their membranes to support energy production and control cytosolic pH. The regulation of intracellular pH in living organisms is a complex process that involves the interaction of proton (H^+) membrane transport, cytosolic buffering, and cellular reactions that produce and consume H^+. In contrast to that of animal cells, the extracellular pH of plant cells is quite acidic. Plasma membrane–bound H^+–adenosine triphosphatases (H^+-ATPases) create a 10- to 100-fold H^+ gradient. The influx of H^+ that leads to cytosolic acidification is thought to have intracellular signaling functions because H^+ can move and carry out responses within the cell before being neutralized (*1*, *2*). However, it remains unclear how H^+ triggers these functions. Accumulating evidence has shown that variations in cytosolic [H^+] (pH_{cyt}) greatly affect cytosolic Ca^{2+} concentration and Ca^{2+} signaling (*3*–*6*). Correlations have been observed between Ca^{2+} signals and pH dynamics during developmental processes, such as the growth of pollen tubes and root hairs (*7*–*9*). Furthermore, cytosolic Ca^{2+} signals and pH dynamics both reach their peaks simultaneously and exhibit synchronous changes in response to exogenous adenosine triphosphate (ATP), auxin, and mechanical stimulation (*10*). Nevertheless, most approaches used in plants to manipulate intracellular pH rely on chemical clues that are invasive in nature and often act with a delay, making it difficult to localize and separate pH and Ca^{2+} signals in real time. Although the light-driven outward proton pump Archaerhodopsin (ArchT) has proven effective in altering intracellular pH dynamics in mammalian cells (*11*), its application in plant cell pH manipulation is hindered by the need for consistently high light intensity and the relatively slow H^+ transport. As a result, our knowledge about how H^+ affects Ca^{2+} signals in plant cells remains rather limited.

The long-sought natural K^+-conducting channelrhodopsins (KCRs) have been discovered in stramenopile protists. Among them, light-gated *Hc*KCR1 and *Hc*KCR2 from the pseudofungus *Hyphochytrium catenoides* were identified as K^+- and H^+-permeable but Ca^{2+}-impermeable channels (*12*–*14*). Though both *Hc*KCR1 and *Hc*KCR2 can serve as potential optogenetic tools for manipulating cytosolic pH in plant cells, the specific activation spectrum of *Hc*KCR2 makes it more practical when combined with genetically encoded fluorescent indicators (*14*). In this study, we used the blue-shifted *Hc*KCR2 to trigger spatially and temporally defined H^+ signals and to monitor rapid downstream responses in *Arabidopsis* guard cells. Thereby, we uncovered a cause-and-effect relationship between H^+ and Ca^{2+} release in guard cells.

*Hc*KCR2 operates as a light-gated proton channel in planta

Although *Hc*KCRs have been shown to exhibit a high H^+/K^+ permeability ratio (about 3×10^4), they operate as K^+-conducting, Ca^{2+}-impermeable channels under neutral extracellular pH conditions, which are typical for animal cells (fig. S1, A and B) (*12*–*14*). However, the plant extracellular space is acidic and has a K^+ concentration that ranges from micromolar to submillimolar levels (*15*, *16*). To simulate these plant settings in *Xenopus* oocytes, we reduced the external K^+ concentration to 1 mM and decreased the pH to 3, thus providing equal molar concentrations of K^+ and H^+. Under these conditions, light- and voltage-induced *Hc*KCR2 inw currents reversed direction around 0 mV (fig. S1C). Lowering the H^+ concentration from pH 3 to 6 in the bath shifted the reversal potential by more than −120 mV (fig. S1, C and D). This indicates that *Hc*KCR2 inward currents—under plant-like physiological conditions—are carried by H^+. This characteristic makes *Hc*KCR2 a useful tool for studying the causality of H^+ influx and cytosolic pH changes in plant Ca^{2+} signaling.

To test whether *Hc*KCR2 is functional in plant cells, an enhanced yellow fluorescent protein (eYFP)–tagged *Hc*KCR2 or control construct was transiently expressed in *Nicotiana benthamiana* leaves. After 48 hours, eYFP fluorescence was detected at the plasma membrane of *Hc*KCR2-expressing mesophyll protoplasts derived by enzymatic digestion of leaf samples (fig. S2A). With the patch-clamp technique in the whole-cell configuration, blue-light stimulation provoked macroscopic currents in *Hc*KCR2-expressing mesophyll protoplasts but not in controls (fig. S2, B to D), indicating proper targeting and functional activity of this channel in plant cells.

*Hc*KCR2 induces cytosolic H^+ and Ca^{2+} transients in guard cells

To use the light-gated H^+ channel *Hc*KCR2 as a tool to noninvasively manipulate the cytosolic pH of plant cells, we stably coexpressed the channel along with the genetically encoded pH reporter pHuji (*17*) in the model plant *Arabidopsis*. *Arabidopsis* plants expressing *Hc*KCR2 were grown under red-light conditions and exhibited a similar phenotype as control plants or plants expressing *Guillardia theta* anion channelrhodopsin 1 (*Gt*ACR1) (fig. S3). We focused our *Hc*KCR2 study on stomatal guard cells because the regulation of plasma membrane ion channels in response to pH and Ca^{2+} signals is known to a great extent in this cell type (*3*–*6*). After 0.1-, 1.0-, and 5-s light pulses, we observed transient changes in pHuji fluorescence, which reflects pronounced cytosolic acidification of guard cells (Fig. 1, A and B, and movie S1). The amplitude of the pH reporter signal was saturated at approximately 1 mW/mm^2 (fig. S4A) and was proportional to the duration of the light pulse (Fig. 1, B and D). We estimated the *Hc*KCR2-triggered cytosolic pH changes based on an in vivo calibration of the pHuji sensor (fig. S5). Stimulations as short as 1 s provoked pH transients on the order of 0.2 pH units that lasted for more than 100 s (Fig. 1B).

To address the long-standing question of how H^+ affects Ca^{2+} signals in plant cells, we tested if the light-imposed pH changes could provoke cytoplasmic Ca^{2+} signals by coexpressing *Hc*KCR2 with the Ca^{2+} reporter R-GECO1 (*18*). After light stimulation, a cytoplasmic H^+ burst triggered biphasic Ca^{2+} signals in guard cells expressing *Hc*KCR2 but not in control cells

[1]Molecular Plant Physiology and Biophysics, Julius-von-Sachs Institute for Biosciences, Biocenter, Würzburg University, Julius-von-Sachs-Platz 2, D-97082 Würzburg, Germany. [2]College of Life Sciences, State Key Laboratory of Crop Genetics and Germplasm Enhancement and Utilization, Nanjing Agricultural University, Nanjing 210095, China.
*Corresponding author. Email: shouguang.huang@uni-wuerzburg.de (S.H.); rainer.hedrich@uni-wuerzburg.de (R.H.)

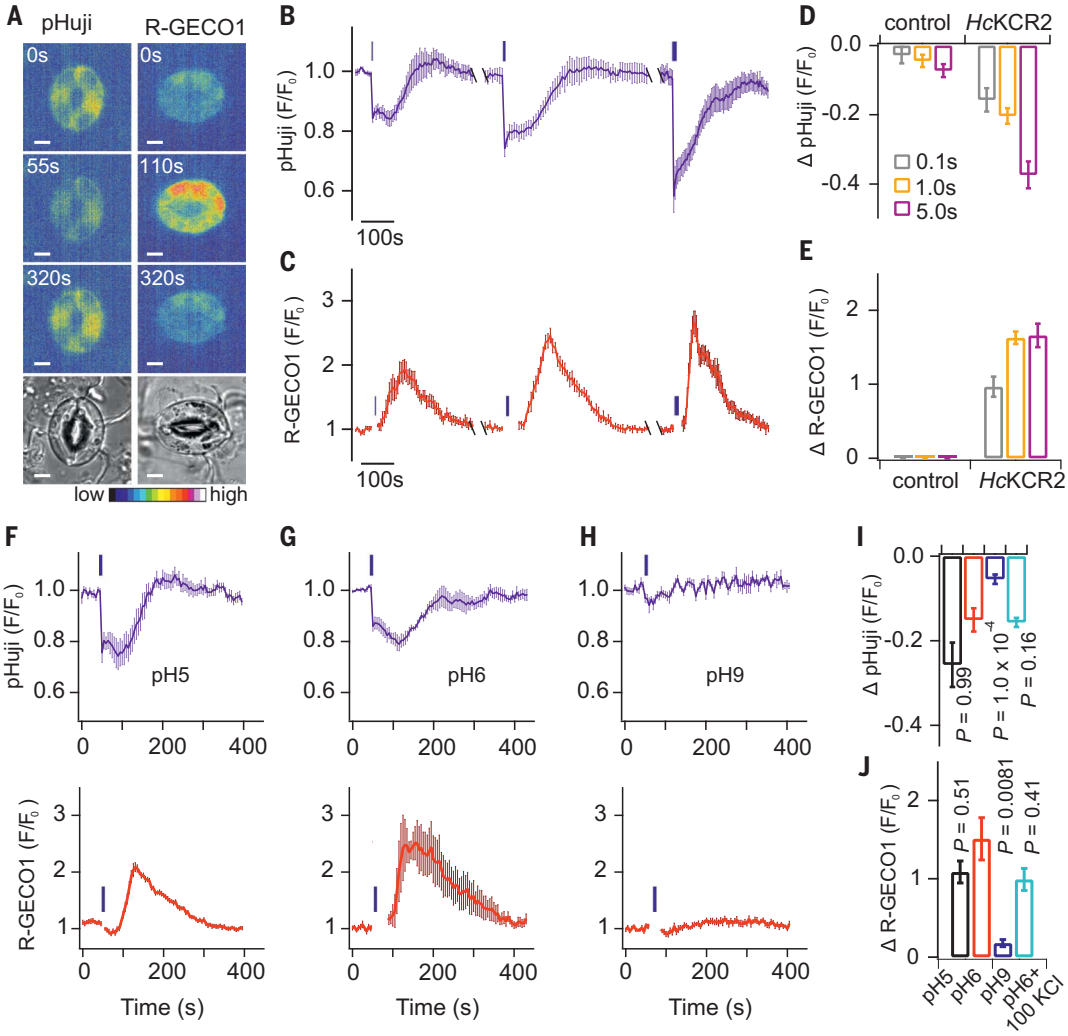


Fig. 1. Cytosolic acidification triggers a Ca^{2+} transient in guard cells. (**A**) Pseudo-color images indicating the cytosolic pH (left column) and Ca^{2+} concentration (right column) in *Hc*KCR2-expressing *Arabidopsis* guard cells. During the experiment, the cells were stimulated with a 1-s blue-light pulse at 55 s. Scale bars are 5 μm. See also movies S1 and S2. (**B** and **C**) Time-dependent changes in pHuji (B) or R-GECO1 (C) signals in *Hc*KCR2-expressing guard cells in response to 0.1-, 1-, and 5-s blue-light pulses (indicated by the increasing thickness of the blue bars above the traces). Note that the R-GECO1 curves are interrupted at the time period of blue-light stimulation. The double slashes on the graph represent 10 min. Error bars show SEM. (n = 9 stomata for pHuji, n = 8 stomata for R-GECO1). F/F_0, ratio of time-dependent fluorescence to fluorescence at the start of the experiment. (**D** and **E**) Average change in pHuji (D) and R-GECO1 (E) signals in *Hc*KCR2-expressing guard cells provoked by blue-light pulses of 0.1, 1, and 5 s. Error bars show SEM (n = 9 to 12 stomata for pHuji, and n = 9 or 10 stomata for R-GECO1). (**F** to **H**) pHuji (top) and R-GECO1 (bottom) signals recorded in *Hc*KCR2-expressing guard cells at extracellular pH values of 5 (F), 6 (G), and 9 (H). The cells were stimulated with a 1-s blue-light pulse (blue bars). The curves of the R-GECO1 signals are interrupted during the time period of blue-light stimulation. Error bars show SEM (n = 6 to 10 stomata for pHuji, and n = 5 to 12 stomata for R-GECO1). (**I** and **J**) Average change in pHuji (I) and R-GECO1 (J) signals in *Hc*KCR2-expressing guard cells at extracellular pH 5, 6, and 9 and at pH 6 with 100 mM KCl (as indicated by the x axis). Error bars show SEM (n = 6 to 10 stomata for pHuji, and n = 5 to 12 stomata for R-GECO1).

(fig. S6A). We propose that the initial fast Ca^{2+} signal is likely triggered by a fast membrane potential swing because the anion-conducting channelrhodopsin *Gt*ACR1 (*19*) also produced a comparable initial Ca^{2+} response but did not elicit a delayed, poststimulation Ca^{2+} signal typical of *Hc*KCR2 (fig. S6B). We thus focused on this *Hc*KCR2-specific, delayed Ca^{2+} signal and studied its interrelation with the H^+ signal. In *Hc*KCR2-expressing guard cells, the delayed transient rise in Ca^{2+} had a lag time of approximately 30 s relative to the cytosolic acidification (Fig. 1, B and C). Specifically, H^+ and Ca^{2+} transients reached their peak amplitudes after 1 and 60 s upon 0.1 s of blue-light stimulation, respectively (Fig. 1, B and C, and movies S1 and S2). This behavior indicates that guard cells operate an H^+-induced Ca^{2+} release system. When the duration of the blue-light pulse was increased from 0.1 to 1 or 5 s, cytoplasmic acidification increased in proportion to the duration of the light pulse (Fig. 1, B and D). Additionally, the initial slope of the Ca^{2+} rise became faster (Fig. 1C and fig. S7). By contrast, an increase in the amplitude of the Ca^{2+} signal was observed only when the blue-light duration was extended from 0.1 to 1 s, whereas prolonging the duration to 5 s did not further increase the Ca^{2+} signal (Fig. 1, C and E).

To test for H^+-induced Ca^{2+} release in guard cells, we increased the pH from 5 to 6 or 9 (Fig. 1, F to H). As a result, the amplitude of the H^+ transient followed the changes in H^+ gradient. However, the peak of the Ca^{2+} transients remained unchanged when recorded at pH 5 or 6 (Fig. 1, F and G). This resembles the all-or-nothing Ca^{2+} signal in response to 1 to 5 s of blue-light stimulation (Fig. 1, C to E). When guard cells were subjected to *Hc*KCR2 stimulation at pH 9, neither a H^+ nor a Ca^{2+} transient was observed (Fig. 1, H to J), indicating that the H^+ influx is essential and sufficient to provoke the Ca^{2+} transient. This conclusion is further corroborated by the observation that a 100-fold increase in external K^+ at pH 6 has no significant effect on the H^+-induced Ca^{2+} transient (Fig. 1I and fig. S8).

H^+ influx triggers Ca^{2+} release from the endoplasmic reticulum in guard cells

Having shown that *Hc*KCR2-induced pH transients rely on the H^+-rich extracellular matrix, we next asked whether the *Hc*KCR2-induced Ca^{2+} transient is of extracellular or endogenous origin. To explore the first possibility, we added 1 mM of the Ca^{2+} chelator EGTA to the extracellular medium. When compared with use of the standard media that contained 1 mM Ca^{2+}, use of EGTA modified media resulted in depletion of the inward Ca^{2+} gradient, which affected the Ca^{2+} amplitude by about 20% (Fig. 2, A, B, D, and E). Considering the observed delay in the H^+-induced Ca^{2+} response (Fig. 1, B and C), we speculated that H^+ ions potentially traverse a short distance to target an endogenous Ca^{2+} store. We thus tested if the endoplasmic reticulum (ER) is the potential Ca^{2+} store that is modulated

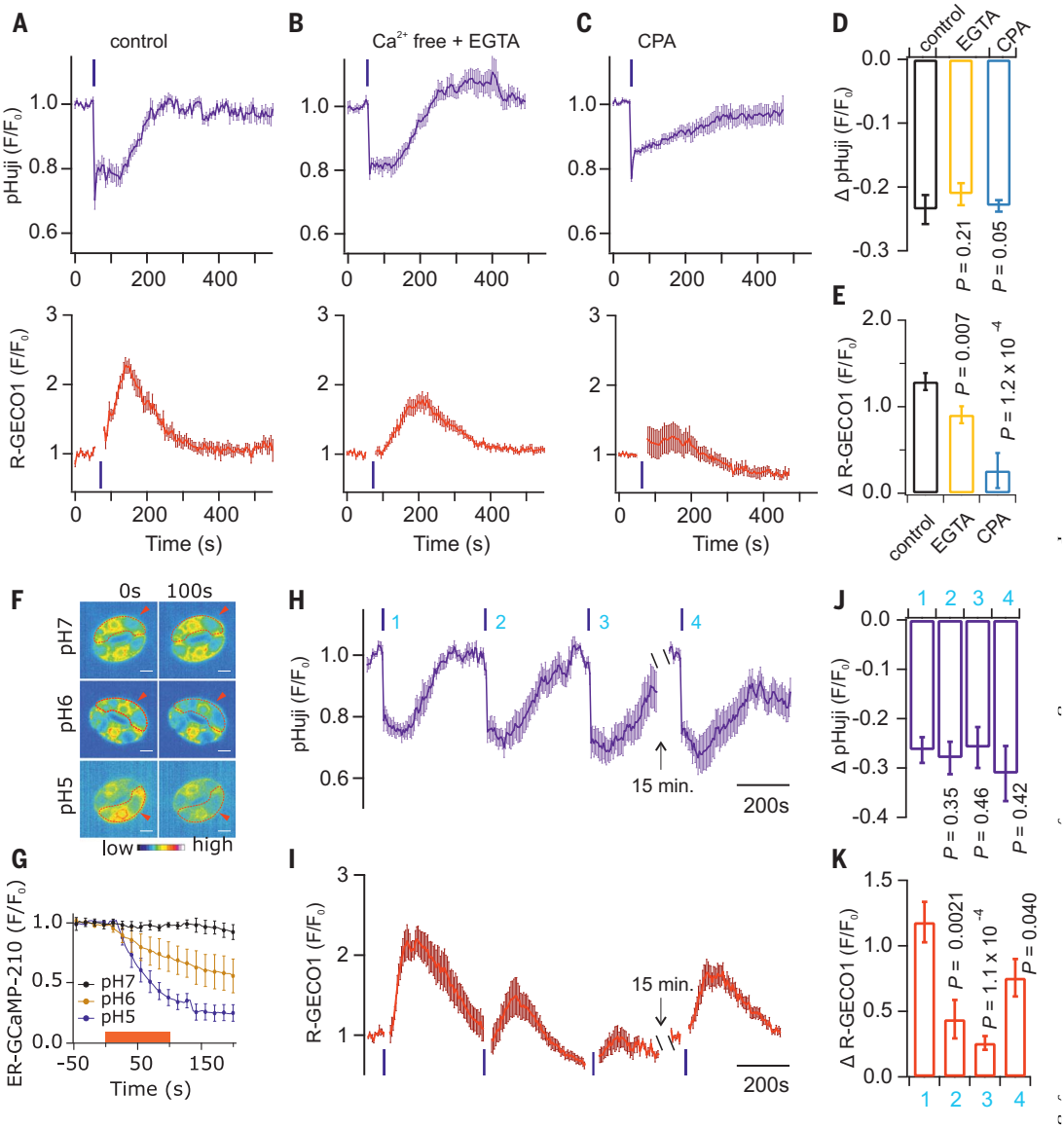


Fig. 2. Cytosolic acidification releases ER Ca^{2+} in guard cells. (**A** to **C**) Impact of pharmaceutical treatment on blue light–induced pHuji (top) or R-GECO1 (bottom) signals in *Hc*KCR2-expressing guard cells. Cells were recorded in control bath solution (A), in a solution with no Ca^{2+} but with 1 mM EGTA (B), and in a solution with 20 μM CPA (C). Each treatment was followed by illumination with a 1-s blue-light pulse (blue bars). The curves of the R-GECO1 signals are interrupted during the time period of blue-light stimulation. (**D** and **E**) Average change in pHuji (D) or R-GECO1 (E) signals in *Hc*KCR2-expressing guard cells exposed to control solution, a solution without Ca^{2+} but with EGTA, or a solution with CPA (as indicated by the *x* axis). Error bars show SEM. (n = 7 to 9 stomata for pHuji, and n = 9 to 13 stomata for R-GECO1). *P* values were determined with a Student's *t* test. (**F**) Current injection of H^+ into guard cells expressing calcium sensor ER-GCaMP-210. False-color images depict the response of guard cells to a 100-s voltage pulse, which changes the membrane potential from −100 to −20 mV, resulting in the injection of H^+ into the cytosol. The tips of one of the capillaries in a double-barreled electrode were filled with pH buffers at pH 7 (control), 6, and 5. Red triangles indicate the guard cells loaded with H^+. Scale bars are 5 μm. (**G**) The time course of the change in ER-GCaMP-210 signal in response to current injection of H^+, as described in (F). The orange bar indicates the period during which H^+ was current-injected. Error bars show SEM. (n = 5 guard cells for pH 7, and n = 4 guard cells for pH 6 and 5). (**H** and **I**) Time course of average pHuji (H) and R-GECO1 (I) signals in *Hc*KCR2-expressing guard cells exposed to repetitive 1-s blue-light pulses (blue bars). The double slash indicates a 15-min interval. The curve of the R-GECO1 signal is interrupted during the time period of blue-light stimulation. (**J** and **K**) Average change in pHuji (J) and R-GECO1 (K) signals in *Hc*KCR2-expressing guard cells, induced by successive blue-light pulses; the pulse number is indicated on the graph. Error bars show SEM (n = 8 stomata for pHuji, and n = 9 stomata for R-GECO1). *P* values were determined with a Student's *t* test.

by *Hc*KCR2. When guard cells were challenged with 20 μM cyclopiazonic acid (CPA), an inhibitor of the ER Ca^{2+}-ATPase isoforms (*20*), the ER Ca^{2+} pool was slowly depleted. However, cytosolic Ca^{2+} concentration remained unchanged (fig. S9), which might be attributed to the smaller size of the guard cells' cytoplasm compared with that of root cells (*21*) and, as a result, a relative higher activity of guard cell plasma membrane Ca^{2+}-ATPases. Next, we studied the impact of CPA on blue light–induced Ca^{2+} transients in *Hc*KCR2-expressing guard cells and found that the Ca^{2+} signal was suppressed (to 17%) compared with that in control cells, whereas the pH signal remained unaffected (Fig. 2, C to E, and fig. S10).

To test if *Hc*KCR2-mediated cytosolic acidification indeed initiates Ca^{2+} release from the ER, we equilibrated the guard cells' cytosolic pH with external ammonium acetate and iontophoretically loaded H^+ into the guard cells' cytosol. In both cases, we observed H^+-induced Ca^{2+} release from the ER (Fig. 2, F and G, and fig. S11). Together, our findings indicate that *Hc*KCR2 targets a Ca^{2+} store that is pH and CPA sensitive.

Given that Ca^{2+} refilling of the ER by ATPases is slower than channel-mediated Ca^{2+} release, we assumed that repetitive H^+-induced Ca^{2+} release through *Hc*KCR2 stimulation will eventually lead to Ca^{2+} depletion. We challenged this hypothesis by subjecting *Hc*KCR2-expressing guard cells to repetitive blue-light pulses. Consecutive stimulation with 1-s blue-light pulses every 6 min did not affect the amplitude of the H^+ transient but caused a stepwise drop in the Ca^{2+} signals (Fig. 2, H to J). After a 15-min recovery period, however, another blue-light stimulation again triggered a strong Ca^{2+} signal (Fig. 2, I and K). Taken together, our guard cell data suggest that H^+ influx triggers Ca^{2+} release from the ER store, leading to store depletion.

*Hc*KCR2-assisted guard cell excitation allows remote control of stomatal movement

To answer the question of how the H^+-induced Ca^{2+} transient contributes to guard cell action,

we used intracellular electrodes to monitor the effect of *Hc*KCR2 on membrane potential. Blue-light stimulation for 0.1 s in *Hc*KCR2 guard cells led to an instantaneous membrane depolarization from a resting potential of −140 to −35 mV; conversely, control cells remained electrically silent (Fig. 3, A and D). After this initial response, the membrane potential first recovered by approximately 20 mV before entering a second extended depolarization phase lasting more than 70 s, which was not detected in *Gt*ACR1-expressing guard cells (*22*). This demonstrates that *Hc*KCR2 postdepolarization electrical signals are not caused by the initial membrane depolarization but by cytoplasmic H^+ and/or Ca^{2+} transients, which may activate endogenous plasma membrane ion channels. We have previously shown that guard cell cytosolic Ca^{2+} signals activate S-type anion channels SLAC1 and SLAH3 (*6*, *23*). Consequently, we investigated the potential role of H^+-induced Ca^{2+} signals in activating guard cell anion channels. In contrast to wild-type (WT) guard cells, the postdepolarization was completely lacking in guard cells that do not express either SLAC1 or SLAH3 (*slac1/slah3* mutant) (Fig. 3, A and D), supporting the idea that anion channels contribute to the poststimulus depolarization phase.

Guard cells do not share plasmodesmata with their neighbors in the leaf epidermis. Therefore,

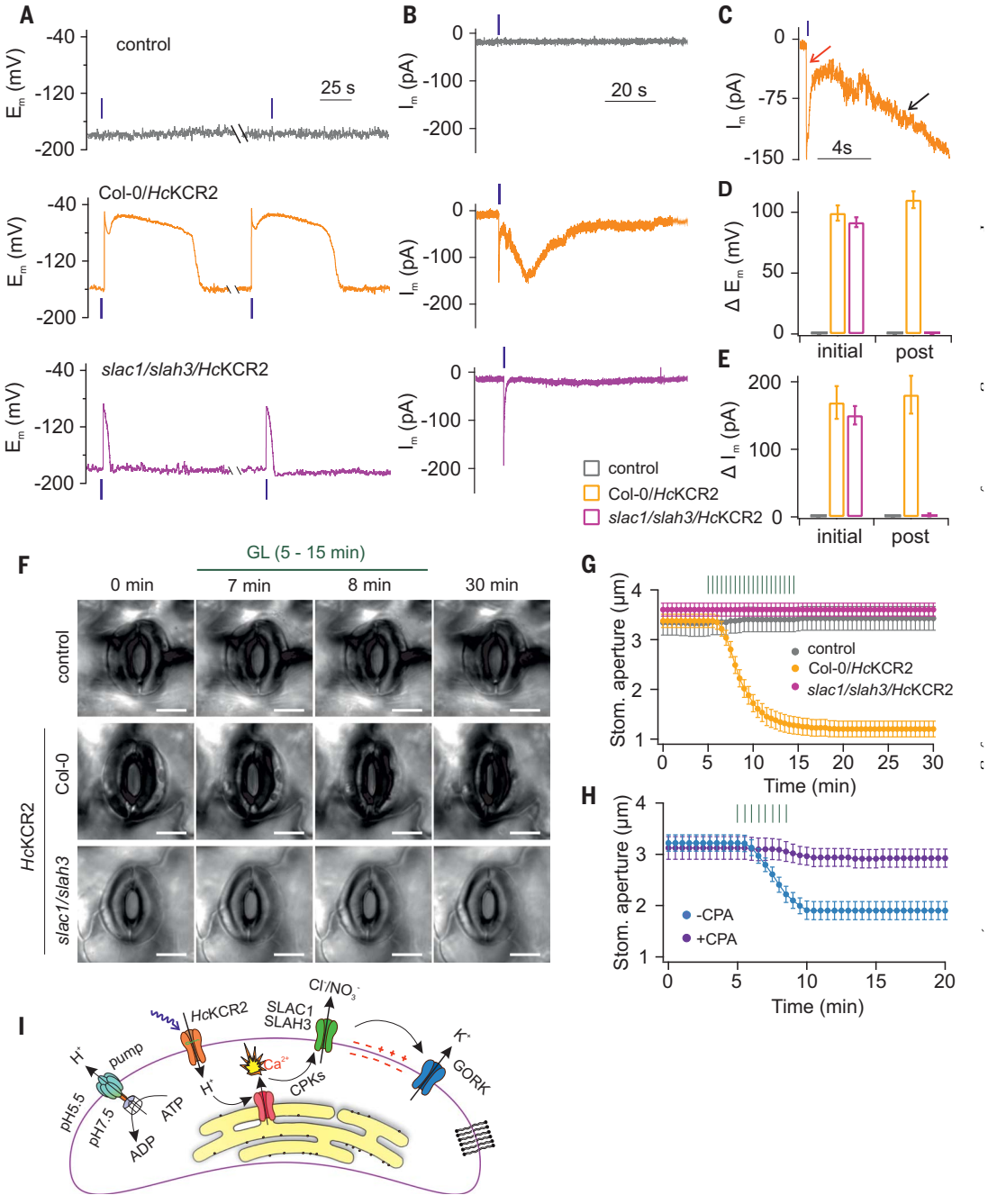


Fig. 3. Activation of guard cell endogenous anion channels by *Hc*KCR2 enables light-controlled stomatal movement. (**A**) Membrane potential response to 0.1-s blue-light pulses measured in guard cells of control (top), *Hc*KCR2-expressing WT (Col-0/*Hc*KCR2; middle), and *Hc*KCR2-expressing *slac1/slah3* mutant (*slac1/slah3*/*Hc*KCR2; bottom) leaves. The double slashes on the *x* axis indicate 8-min intervals. (**B**) Plasma membrane currents evoked by 0.1-s blue light pulses in guard cells of control (top), Col-0/*Hc*KCR2 (middle), and *slac1/slah3*/*Hc*KCR2 (bottom) clamped at −150 mV. E_m, plasma membrane potential; I_m, plasma membrane current. (**C**) Expanded current trace of a Col-0/*Hc*KCR2 guard cell that was stimulated with a 0.1-s blue-light pulse. Red and black arrows indicate the initial fast photocurrent and the slow anion current, respectively. (**D**) Average initial and postdepolarization values for guard cells provoked with 0.1-s blue-light pulses. Error bars show SEM (*n* = 7 to 14 guard cells). (**E**) Mean values of photocurrent (initial) and slow-activating anion currents (post) for guard cells provoked by 0.1-s blue light pulses. Error bars show SEM (*n* = 5 or 6 guard cells). (**F**) Representative images of control, Col-0/*Hc*KCR2, and *slac1/slah3*/*Hc*KCR2 stomata, obtained before, during, and after a train of green-light (GL) pulses (duration: 0.5 s; interval: 30 s; power density: 0.16 mW/mm^2), as depicted above the traces in (G). See also movie S3. (**G**) Time course of stomatal movement in control, Col-0/*Hc*KCR2, and *slac1/slah3*/*Hc*KCR2 leaves, which were stimulated with a train of green-light pulses, as indicated in the graph. Error bars show SEM. (**H**) CPA inhibits green light–induced stomatal closure in *Hc*KCR2-expressing plants. Eight pulses of green light (duration: 0.5 s; interval: 30 s; power density: 0.16 mW/mm^2) were applied to the leaf discs treated with or without 100 μM CPA. Error bars show SEM. (**I**) A schematic model for H^+-induced Ca^{2+} release in guard cells. Blue light activates *Hc*KCR2, triggering H^+-induced Ca^{2+} release from the ER, which sequentially leads to the activation of anion channels SLAC1 and SLAH3 and guard cell outward-rectifying K^+ channel (GORK). Consequently, anion and K^+ efflux causes guard cell shrinkage and stomatal pore closure. ADP, adenosine diphosphate.

ion currents can be recorded in individual guard cells in the voltage-clamp configuration (*24*, *25*). The blue light–induced *Hc*KCR2-dependent depolarization event in guard cells is mirrored by an underlying ion current profile that is uncovered under voltage-clamp conditions. Light stimulation of single guard cells provoked a transient inward photocurrent mediated by *Hc*KCR2-dependent H^+ influx (Fig. 3, B and C). Magnification of early events documents that photocurrent (H^+ current or H^+ spark) merged with an emerging S-type anion current transient, which reached its peak after 10 s and then decayed within 70 s (Fig. 3C). In line with membrane potential measurements, blue light induced transient photocurrents but not anion currents in the *slac1/slah3* mutant (Fig. 3, B and E).

Because SLAC1 and SLAH3 are targets of H^+-induced Ca^{2+} transients and are key to prolonged guard cell depolarization (Fig. 3, A to E), we tested the potential of *Hc*KCR2 as a tool for remote-controlling stomatal movement. To investigate stomatal action, we used green light to activate *Hc*KCR2 because guard cells are less sensitive to green light than to blue light (*26*). When intact *Arabidopsis* leaves were challenged by 0.5-s green-light pulses every 30 s, the stomata of control plants remained open throughout the course of the experiment, whereas in *Hc*KCR2-expressing WT plants, stomatal closing was initiated in 2 min and completed in less than 10 min (half-time 4.5 min) (Fig. 3, F and G, and movie S3). *Hc*KCR2 stimulation in the *slac1/slah3* mutant, however, did not result in stomatal closure (Fig. 3, F and G). Given that CPA inhibition of ER Ca^{2+} pumps prevented *Hc*KCR2-mediated Ca^{2+} release from endogenous stores (Fig. 2C) and suppressed action potentials and Ca^{2+} signaling in the Venus flytrap (*27*), we finally asked if CPA blocks stomatal action. Indeed, when *Hc*KCR2 WT *Arabidopsis* leaves were pretreated with CPA, green-light pulses did not induce stomatal closure (Fig. 3H). Taken together, these findings indicate that *Hc*KCR2-mediated stomatal closure involves Ca^{2+} release from CPA- and H^+-sensitive ER Ca^{2+} stores, which activate the SLAC1 and SLAH3 anion channels.

Discussion

Busa and Nuccitelli highlighted the interdependence of intracellular pH and Ca^{2+} concentration in various cell types (*28*). However, subsequent research in the animal field has yielded contradictory findings, with some authors suggesting that an increase in cytoplasmic Ca^{2+} is caused by a reduction in cellular pH and others proposing that it is due to an increase in pH (*1*). Similarly, studies in plants have also produced inconsistent results regarding pH-dependent Ca^{2+} signaling (*3–5*, *9*, *10*, *29*). In this study, we optically controlled plasma membrane H^+ influx with *Hc*KCR2 and elucidated a H^+-induced Ca^{2+} release mechanism from the ER in guard cells. Now, a fundamental question remains: How does cytosolic acidification lead to an all-or-nothing Ca^{2+} transient? Although the mechanism of ER Ca^{2+} sourcing in response to cytoplasmic acidification in plant cells remains elusive, studies in animal systems have documented the impact of pH on spontaneous Ca^{2+} release from intracellular Ca^{2+} stores (*30*, *31*). Bax inhibitor-1 (BI-1), which is present in both animal and plant ER, serves as a pH-sensitive Ca^{2+} channel. A neutral pH (pH 7) facilitates the channel opening, whereas both lower and higher pH values result in channel closure (*32–34*). In plant cells, activation of *Hc*KCR2 by blue light causes a drop in pH of the cytosol, which may spark Ca^{2+} release from the ER (Fig. 3I). On the basis of these features, BI-1 emerges as a promising candidate for future investigations aimed at identifying the H^+-sensitive ER Ca^{2+} channel. Further, the recently identified ER-localized Ca^{2+} channel WeiTsing (WTS) could also be a potential candidate (*35*), although its pH sensitivity has yet to be examined. We cannot rule out that *Hc*KCR2-mediated cytosolic pH changes may influence H^+ or cation Ca^{2+} exchanger transport activity on the ER membrane (*10*, *36*). However, considering the rapid onset of the cytosolic Ca^{2+} transient (fig. S7), we favor the involvement of a H^+-sensitive, Ca^{2+}-permeable, ER-located ion channel.

Our data unequivocally demonstrate that *Hc*KCR2 serves as a tool for remote-controlling H^+ and Ca^{2+} signaling and suggest a causal relationship between H^+ and Ca^{2+} signaling in guard cells. Finally, by leveraging the distinctive properties of specific channelrhodopsins, we may broaden our understanding about the resilience of plants toward changing environmental conditions.

REFERENCES AND NOTES

1. G. Molinari, E. Nervo, *Biochem. J.* **478**, 895–910 (2021).
2. A. S. Raghavendra, W. Ye, T. Kinoshita, *Front. Plant Sci.* **14**, 1148689 (2023).
3. H. R. Irving, C. A. Gehring, R. W. Parish, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1790–1794 (1992).
4. M. M. Islam *et al.*, *Plant Cell Physiol.* **51**, 1721–1730 (2010).
5. X. Zhang, F. C. Dong, J. F. Gao, C. P. Song, *Cell Res.* **11**, 37–43 (2001).
6. S. Huang *et al.*, *New Phytol.* **224**, 177–187 (2019).
7. E. Michard, A. A. Simon, B. Tavares, M. M. Wudick, J. A. Feijó, *Plant Physiol.* **173**, 91–111 (2017).
8. G. B. Monshausen, T. N. Bibikova, M. A. Messerli, C. Shi, S. Gilroy, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20996–21001 (2007).
9. K. Li *et al.*, *New Phytol.* **230**, 2292–2310 (2021).
10. S. Behera *et al.*, *Plant Cell* **30**, 2704–2719 (2018).
11. C. E. T. Donahue, M. D. Siroky, K. A. White, *J. Am. Chem. Soc.* **143**, 18877–18887 (2021).
12. J. Vierock *et al.*, *Sci. Adv.* **8**, eadd7729 (2022).
13. S. Tajima *et al.*, *Cell* **186**, 4325–4344.e26 (2023).
14. E. G. Govorunova *et al.*, *Nat. Neurosci.* **25**, 967–974 (2022).
15. R. Hedrich, *Physiol. Rev.* **92**, 1777–1811 (2012).
16. J. I. Schroeder, J. M. Ward, W. Gassmann, *Annu. Rev. Biophys. Biomol. Struct.* **23**, 441–471 (1994).
17. Y. Shen, M. Rosendale, R. E. Campbell, D. Perrais, *J. Cell Biol.* **207**, 419–432 (2014).
18. Y. Zhao *et al.*, *Science* **333**, 1888–1891 (2011).
19. E. G. Govorunova, O. A. Sineshchekov, R. Janz, X. Liu, J. L. Spudich, *Science* **349**, 647–650 (2015).
20. F. Liang, H. Sze, *Plant Physiol.* **118**, 817–825 (1998).
21. F. Resentini *et al.*, *Plant Physiol.* **187**, 603–617 (2021).
22. S. Huang *et al.*, *Sci. Adv.* **7**, eabg4619 (2021).
23. R. Hedrich, D. Geiger, *New Phytol.* **216**, 46–61 (2017).
24. A. Grabov, M. R. Blatt, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4778–4783 (1998).
25. M. R. Roelfsema, R. Steinmeyer, M. Staal, R. Hedrich, *Plant J.* **26**, 1–13 (2001).
26. J. J. Jones, S. Huang, R. Hedrich, C.-M. Geilfus, M. R. G. Roelfsema, *New Phytol.* **236**, 1237–1244 (2022).
27. S. Scherzer *et al.*, *Curr. Biol.* **32**, 4255–4263.e5 (2022).
28. W. B. Busa, R. Nuccitelli, *Am. J. Physiol.* **246**, R409–R438 (1984).
29. S. Wagner *et al.*, *New Phytol.* **224**, 1668–1684 (2019).
30. C. H. Orchard, *J. Physiol.* **384**, 431–449 (1987).
31. H.-R. Kim *et al.*, *J. Biol. Chem.* **283**, 15946–15955 (2008).
32. G. Guo *et al.*, *Structure* **27**, 1013–1021.e3 (2019).
33. Y. Chang *et al.*, *Science* **344**, 1131–1135 (2014).
34. G. Bultynck, S. Kiviluoto, A. Methner, *Sci. Signal.* **7**, pe22 (2014).
35. W. Wang *et al.*, *Cell* **186**, 2656–2671.e18 (2023).
36. M. Corso, F. G. Doccula, J. R. F. de Melo, A. Costa, N. Verbruggen, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3966–3971 (2018).

ACKNOWLEDGMENTS

Funding: This work was funded by German Research Foundation (DFG) Reinhart Koselleck Grant 415282803 (R.H.) and China Scholarship Council no. 202006855028 (L.S.). **Author contributions:** Conceptualization: S.H., R.H.; Methodology: S.H., L.S.; Investigation: S.H., L.S., M.R.G.R., D.B.; Visualization: S.H., L.S., M.R.G.R., D.B., R.H.; Funding acquisition: R.H.; Project administration: R.H.; Supervision: M.R.G.R., D.B., R.H.; Writing – original draft: S.H., R.H.; Writing – review and editing: S.H., L.S., M.R.G.R., D.B., R.H. **Competing interests:** The authors have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. There are no restrictions on materials sharing. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj9696
Materials and Methods
Figs. S1 to S11
Table S1
References (*37–45*)
MDAR Reproducibility Checklist
Movies S1 to S3

Submitted 26 July 2023; accepted 10 November 2023
10.1126/science.adj9696

Post date 15 December 2

Erratum

Erratum for the Research Article "Agriculture and hot temperatures interactively erode the nest success of habitat generalist birds across the United States" by K. S. Lauck *et al.*

In the Research Article "Agriculture and hot temperatures interactively erode the nest success of habitat generalist birds across the United States," the effect sizes were mistakenly calculated using the extremes of the maximum temperature anomaly distribution. The effect sizes are now correctly calculated using 2 SDs above and below the maximum temperature anomaly mean, as described in the main text and methods. The correction does not change the paper's conclusions.

10.1126/science.adn2847

WORKING LIFE

By **Adrian Beckert**

A challenging move

"How do you like your new postdoc?" It was a seemingly innocuous question from a former coworker I met at a conference. But answering it honestly was a challenge. The reality was that 5 months into my postdoc, I was feeling increasingly despairing as my family faced obstacles adapting to life in a new country and I struggled to get my project off the ground. I ended up responding evasively, briefly mentioning a few positive aspects of my new position. But the question was a turning point: I realized my situation wasn't sustainable and that I needed to figure out better ways to cope with the aftermath of our international move.

I had done my Ph.D. in my home country of Switzerland. Toward the end of my program, I told my wife that if I wanted to pursue an academic career, we would probably have to move abroad. She was enthusiastic. She could do her job remotely and had always wanted to live in another country. We also liked the thought of giving our two young daughters an opportunity to be immersed in another language. So, when I was offered a dream postdoc in the United States, we were all excited about this new chapter in our lives.

We expected a big change, but not the seismic shift that met us upon our arrival. We had to move between temporary accommodations before we found a more permanent home, all while rebuilding a household's worth of items from scratch. We hadn't visited the U.S. before, so our family struggled with the cultural transition and disappearance of our usual Swiss routines—letting our kids walk to school on their own and play in the backyard unsupervised. Our cherished support network of friends and family were also an ocean away, and hiring a babysitter wasn't viable because of budget constraints and our kids' aversion to another unfamiliar face.

"We expected a big change, but not the seismic shift that met us upon our arrival."

The stress bled into my work life. I found that I couldn't put in the number of hours I wanted to. This was a problem because I had to learn new measurement techniques for my postdoc and build an experimental setup. I also spent a significant amount of time working on unfinished projects from my Ph.D., which was demanding and left me feeling guilty for not devoting myself fully to my postdoc.

I was determined to succeed at work and at home. So, I pushed myself harder and harder, rarely taking the time I needed to recharge. This overly ambitious plan failed and my perspective darkened, leaving me exhausted and feeling like a failure.

After finding I couldn't honestly tell my former colleague that I was having a positive postdoc experience, I talked to a counselor to get to the bottom of my struggles. I learned that I shouldn't be so hard on myself; moving across continents with a family and starting a new project were inevitably going to be difficult. That helped me view my work more objectively and not compare myself with others in my lab, who seemed to be making faster progress than I. To convince me to go easier on myself, the counselor advised jokingly, "Be more like Trump," alluding to the former U.S. president's tendency to deflect blame away from himself.

My counselor also encouraged me to share my struggles with my supervisor. I set up a meeting and revealed my guilt over sometimes prioritizing old projects and my frustration with the slow progress of my postdoc work. He was empathetic, which relieved a lot of pressure. He told me many postdocs encounter similar challenges and offered me an extended time off to recharge my batteries.

At home, my wife and I also worked on our social network to battle isolation. I joined university workshops and my wife took English lessons that allowed us to expand our social circles. We both scheduled regular video calls with our friends back home to reconnect and share struggles. And we transformed our family time to focus more on fun activities such as traveling and camping.

Juggling work and my home life is still challenging at times, as is living in a new country. But I'm in a better place now. As I continue to push forward in my career, my family and I will surely encounter other tough transitions. But next time, we'll be better equipped, with more strategies to deal with the challenges. ■

Adrian Beckert is a postdoc at the California Institute of Technology.